

# Accuracy of the malaria epidemiological surveillance system data in the state of Amazonas

Alexandre WIEFELS<sup>1</sup>, Bruna Wolfarth-COUTO<sup>2</sup>, Naziano FILIZOLA<sup>3</sup>, Laurent DURIEUX<sup>1,4</sup>, Morgan MANGEAS<sup>1,4</sup>

1 Unité Mixte de Recherche Espace pour le Développement, Montpellier, France.

2 Instituto Nacional de Pesquisas da Amazônia, Manaus, Brasil.

3 Universidade Federal do Amazonas, Laboratório de Potamologia Amazônica, Manaus, Brasil.

4 Institut de Recherche pour le Développement, Montpellier, France.

\* Corresponding author: [alexandre.wiefels@teledetection.fr](mailto:alexandre.wiefels@teledetection.fr).

## ABSTRACT

The Epidemiological Surveillance System for Malaria (SIVEP-Malaria) is the Brazilian governmental program that registers all information about compulsory reporting of detected cases of malaria by all medical units and medical practitioners. The objective of this study is to point out the main sources of errors in the SIVEP-Malaria database by applying a data cleaning method to assist researchers about the best way to use it and to report the problems to authorities. The aim of this study was to assess the quality of the data collected by the surveillance system and its accuracy. The SIVEP-Malaria data base used was for the state of Amazonas, Brazil, with data collected from 2003 to 2014. A data cleaning method was applied to the database to detect and remove erroneous records. It was observed that the collecting procedure of the database is not homogeneous among the municipalities and over the years. Some of the variables had different data collection periods, missing data, outliers and inconsistencies. Variables depending on the health agents showed a good quality but those that rely on patients were often inaccurate. We showed that a punctilious preprocessing is needed to produce statistically correct data from the SIVEP-Malaria data base. Fine spatial scale and multi-temporal analysis are of particular concern due to the local concentration of uncertainties and the data collecting seasonality observed. This assessment should help to enhance the quality of studies and the monitoring of the use of the SIVEP database.

**KEYWORDS:** Erroneous data, Database, Health surveillance.

# Exatidão dos dados do sistema de vigilância epidemiológica da malária no estado do Amazonas

## RESUMO

O Sistema de Vigilância Epidemiológica de Malária (SIVEP-Malária) é um programa governamental brasileiro que arquiva automaticamente todas as informações sobre casos de malária registrados em todas as unidades de saúde e consultórios médicos. O objetivo deste estudo foi avaliar a qualidade dos dados coletados pelo sistema de vigilância e sua precisão. Foram utilizados os dados do SIVEP-Malária para o estado do Amazonas, Brasil, de 2003 a 2014. Um método de limpeza de dados foi aplicado para detectar e remover registros errôneos. Observamos que a coleta de dados não é homogênea entre os municípios e ao longo dos anos. Algumas variáveis tinham diferentes padrões de coleta, falta de dados, dados discrepantes e inconsistências. Dados que dependem do agente de saúde possuem boa qualidade mas aqueles que dependem dos pacientes são frequentemente imprecisos. Mostramos que um pre-processamento meticuloso é necessário para produzir dados estatisticamente corretos a partir do SIVEP-Malária. Análises em escala espacial detalhada ou multi-temporais são particularmente afetadas devido à concentração local de incertezas e a sazonalidade observada na coleta de dados. Esta avaliação deve auxiliar a melhorar os estudos e monitoramentos que fazem uso dos dados do SIVEP.

**PALAVRAS-CHAVE:** Dados errôneos, Base de dados, Vigilância sanitária.

## INTRODUCTION

The Epidemiological Surveillance System for Malaria (SIVEP-Malaria) is the Brazilian governmental program that registers all information about compulsory reporting of detected cases of malaria by all medical units and medical practitioners. It collects and provides information on the number and distribution of malaria cases and deaths in a territory. It is a critical tool for implementation of governmental malaria control programs (Deane 1988; Ferreira and Silva Nunes 2010; Silva *et al.* 2010). It is needed to determine which areas or population groups are most affected by malaria, so that resources on prevention and control can be targeted to the relevant populations (Martens *et al.* 1999; Almeida *et al.* 2010; Taulil and Daniel-Ribeiro 1998). Northern Brazil is an endemic malaria region (Barata 1995, Oliveira-Ferreira *et al.* 2010) and a growing number of researches make use of the Brazilian epidemiological surveillance system of malaria (SIVEP-Malaria) database.

In 1999, the Ministry of Health launched the Plan for Intensification of Malaria Control Actions [*Plano de Intensificação das Ações de Controle de Malária* (PIACM)] as a response to the World Health Organization (WHO) Roll Back Malaria Program (Sampaio *et al.* 2015). In 2003, in partnership with states and municipalities of the Amazon region, the Brazilian Ministry of Health's Secretariat of Health Surveillance (SVS) created the National Program for Prevention and Control of Malaria [*Programa Nacional de Prevenção e Controle da Malária* (PNCM)] for the prevention and control of the disease, continuing the progress made with the PIACM (Pina-Costa *et al.* 2014). The Malaria's Epidemiological Surveillance System (SIVEP-Malaria) is the main tool used by PNCM to improve the production of information about malaria. From 2003 to 2004, the municipal participation in the SIVEP-Malaria program increased from 60% to 100% in the Amazon region where 99% of the malaria cases in Brazil are concentrated (Brasil 2003; Braz *et al.* 2006). In Brazil, patients suspected of having malaria are compulsorily reported to the Ministry of Health through SIVEP-Malaria (Brasil 2014). The reporting form includes the medical unit data, the patient's personal and clinical data, and also the georeferenced data, generating a total of 43 variables. Thus, the SIVEP-Malaria database allows the spatial and temporal monitoring of epidemics but also serves to assess the coverage of diagnosis and treatment (Oliveira-Ferreira *et al.* 2010). However, research find that the SIVEP information is not utilised well by local managers (Costa *et al.* 2010; Hermes *et al.* 2013; Peiter *et al.* 2013). There are a variety of errors spread over the data, mainly due to mistyping or lack of precision on declarations of patients. In 2013 the Ministry of Health decentralized the managing of data acquisition to

the municipal administration, which has generated different data collection methods.

Our hypothesis is that the raw SIVEP-Malaria data is exploitable for epidemiological studies with scientific credibility after preprocessing it. The objective of this study is to point out the main sources of errors in the SIVEP-Malaria database by applying a data cleaning method to assist researchers about the best way to use it and to report the problems to authorities.

## MATERIALS AND METHODS

The usual analysis approach is to begin with descriptive analyses to explore the data (Schoenbach and Rosamond 2000). This type of prior analysis allows one to identify and intercept errors in order to obtain a valid and consistent data set that is technically correct, thus ready for statistical analysis. The filtering mechanism to obtain this data quality is called data cleaning and it is executed in four steps: detecting missing values, outliers, inconsistencies and finally making the corrections (Broeck *et al.* 2005; Jonge and van der Loo 2013; Hang-Hai and Erhard 2000).

The SIVEP-Malaria data is available on database files kept by the Health System Informatics Department (DATASUS) (Brasil 2015). It contains administrative, personal and laboratorial variables as shown in Table 1. The data banks do not include names or other information that might allow identification of the subjects.

The data cleaning was performed over a selection from the database which included all notified and all infected patients (variables UF\_NOTIF & UF\_INFEC) from the state of Amazonas. This selection allows consideration, at the same time, of the cases registered in the Amazonas health centers and those infected in the Amazonas territory but registered abroad responding to different research objectives. From this selection, a new filter picked only the positive cases (variable RES\_EXAM). Positive cases ensure that one can handle the complete record of all variables, such as clinical ones. Both steps resulted at first in 12,286,851 reported cases and then in 1,801,229 positive cases from 2003 to 2014.

## RESULTS

The data collection for each patient was discontinuous over the time and the variables (Table 2). Some variables from the malaria notification form are systematically filled in, while some others are missing. Finding out where the missing values are is the first step in choosing which variable to use in research for a specific study area.

However, the availability of data is not constant over time. From 2003 to 2014, SIVEP-Malaria improved as a surveillance system. The systematicity of the data production

**Table 1.** SIVEP-Malaria variables code names and definition

Variable	Definition	Variable	Definition	Variable	Definition	Variable	Definition
COD_NOTI	Notification number	DT_NASCI	Birth date	MUN_RESI	Municipality of residence	LOC_INFE	Locality of infection
DT_NOTIF	Notification date	ID_PACIE	Patient age	LOC_RESI	Locality of residence	DT_EXAME	Examination date
TIPO_LAM	Active/passive notification	ID_DIMEA	Age writing format	DT_SINTO	First symptoms date	EXAME	Examination method
UF_NOTIF	State of notification	SEXO	Gender	DT_TRATA	Date of treatment	RES_EXAM	Examination results
MUN_NOTI	Municipality of notification	GESTANTE	Pregnancy length	VIVAX	Patient is under Vivax treatment	QTD_CRUZ	Parasitaemia
COD_UNIN	Health unit of notification	NIV_ESCO	Schooling level	FALCIPARUM	Patient is under Falciparum treatment	QTD_PARA	Parasites by mm <sup>3</sup>
COD_AGEN	Health agent code	RACA	Ethnic group	ID_LVC	Follow-up consultation	HEMOPARASI	Hemoparasites
SEM_NOTI	Notification week	COD_OCUP	Employment	PAIS_INF	Country of infection	EXAMINADOR	Examiner code
DT_DIGIT	Date of digitalization	PAIS_RES	Country of residence	UF_INFEC	State of infection	ESQUEMA	Treatment schedule
DT_ENVLO	Data entering into National database date	UF_RESID	State of residence	MUN_INFE	Municipality of infection	SINTOMAS	Symptoms

Administrative data

Personal data

Laboratorial data

**Table 2.** Data availability per variable from 2003 to 2014 for the state of Amazonas

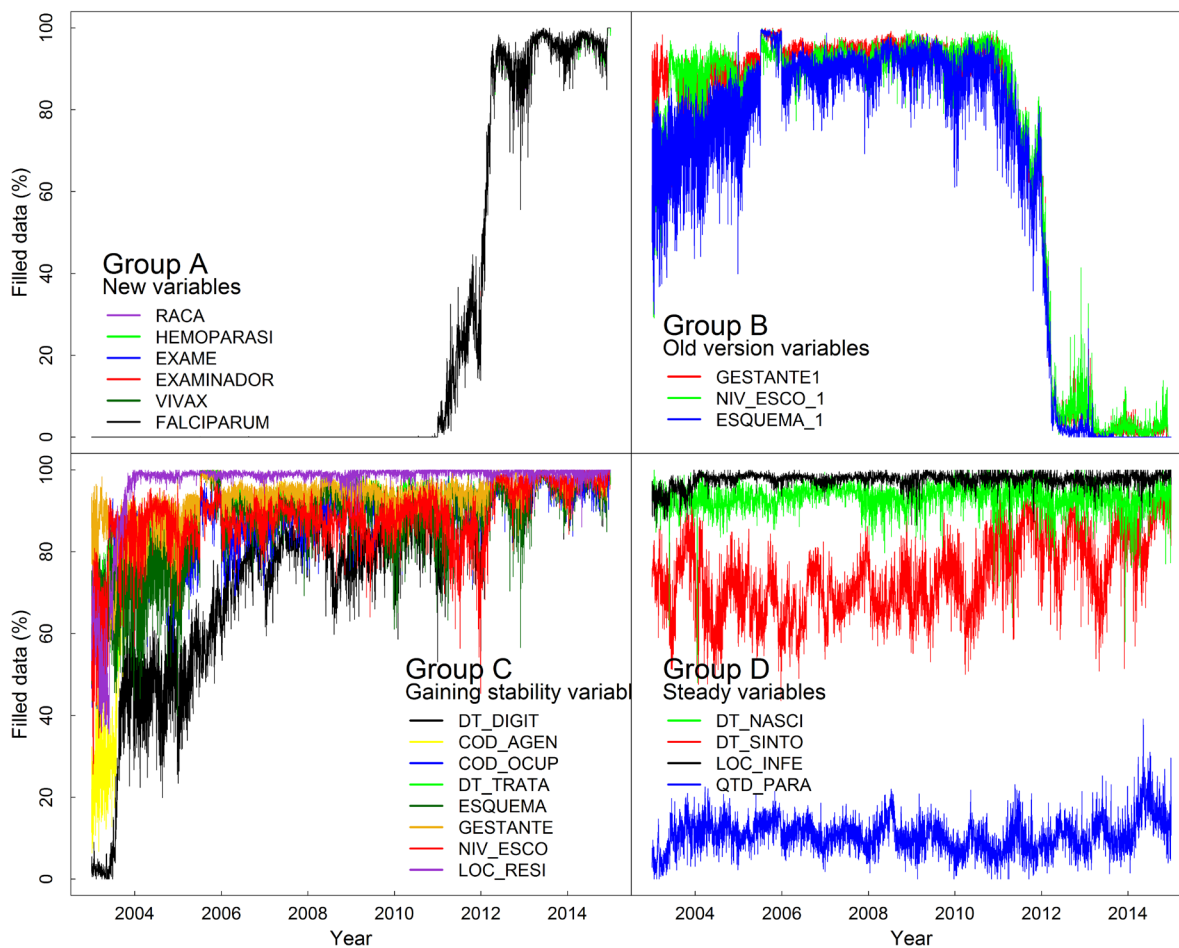
Variable	Data Availability (%)	Variable	Data Availability (%)	Variable	Data Availability (%)
COD_NOTI	100	PAIS_INF	99,90	GESTANTE1	79.50
DT_NOTIF	100	QTD_CRUZ	99,80	NIV_ESCO_1	78.60
DT_ENVLO	100	MUN_RESI	99,80	ESQUEMA_1	75.30
TIPO_LAM	100	UF_RESID	99,30	DT_DIGIT	71.40
UF_NOTIF	100	UF_INFEC	99,30	DT_NASCI	70.90
MUN_NOTI	100	MUN_INFE	99,30	RACA	14.10
COD_UNIN	100	LOC_INFE	97,60	FALCIPARUM	14.10
SINTOMAS	100	LOC_RESI	96,60	VIVAX	14.10
RES_EXAM	100	GESTANTE	93,50	EXAME	14.10
ID_PACIE	99,90	DT_SINTO	93,30	EXAMINADOR	14.10
PAIS_RES	99,90	DT_TRATA	90,50	HEMOPARASI	14.10
ID_LVC	99,90	ESQUEMA	89,40	QTD_PARA	11.90
ID_DIMEA	99,90	COD_AGEN	88,80	SEM_NOTI	0
SEXO	99,90	NIV_ESCO	88,60		
DT_EXAME	99,90	COD_OCUP	88,00		

Variables names are described in Table 1

could be assessed when plotted over time. The entered data values were counted for each variable per day, based on the notification date at the health unit. The proportion between available data, for each variable, and the total of notification collected was plotted (Figure 1).

The results were classified into five *Groups* from *A* to *E*. *Group E* contained the six variables whose values were nearly or always present, and for which a plot was unnecessary. *Group E* variables were: the municipality of infection, the state of residence, the state of infection, the parasitaemia, the municipality of residence, and the country of infection.

Variables of the *Group A* began to be recorded in 2011, and were systematically registered in 2013. By contrast, *Group B*'s three variables developed a systematicity since 2003 but stopped being collected when SIVEP-Malaria's register form changed in 2010. They were transposed to *Group C* with the new method of classification. *Group C* assembled the variables with the most expected behavior, gaining in systematicity and continuous filling over time. *Group D* contained variables that presented a seasonal trend, with a stationary mean value over time. These variables had an annual seasonality. The birth date and the positive malaria cases had a Spearman correlation of -0.65 (Figure 2). It suggested that repeatedly,



**Figure 1.** Classification of the data availability by variable over the years. This figure is in color in the electronic version.

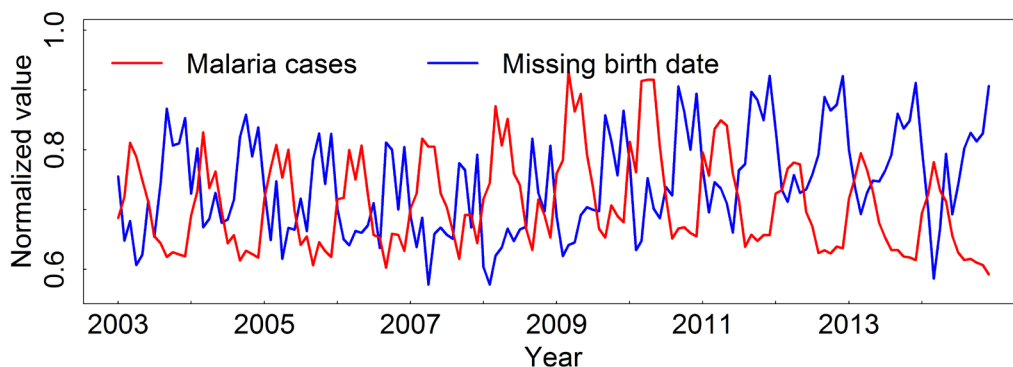
at the epidemiological peak, the health agents made a more efficient data collection, decreasing the amount of missing birth dates on notifications. This periodicity was observed for all the others variables in different proportions but it was more visible on the birth date, particularly because of its stability over time. This observation suggested that the epidemiological cycle affects the notification of other variables.

Most parts of SIVEP-Malaria data are qualitative. Each variable was analyzed separately, comparing the data provided to the expected interval of values to identify unrealistic outliers. There were seven categories representing dates, among which the following present outliers: the birth date - expected to be in the interval between “1900-01-01” and “2014-12-31” - presented 19 occurrences from the year of “1006” to “1900” and 22 from “2030” to “2091”; the symptom date - expected to be between “2000-01-01” and “2014-12-31” - presented 148 occurrences from the year of “1002” to “1999” with a peak around “1200”; and the treatment date - expected to be

between “2003-01-01” and “2014-12-31” – which registered 796 occurrences from the year of “2015” to “9620”.

The exam results (with values from 1 to 11) and the parasitaemia (from 1 to 6) showed 17 times the value “0”, which was not an option, in 2003 for the municipalities of Coari and Manaus. The registration code, with positive values, showed 18 outliers occurrences (16 x “0”, “-1264” and “-682”) from 2003 to 2007. The epidemiological week variable was always absent except for one case in 2003. The municipality of residence has only one unrealistic outlier named “539924” which is not a municipality code. There were 20,242 registered health agents in the SIVEP-Malaria database for Amazonas state. Each one is represented by a code number. From those, 15,692 were responsible for notifications between 2003 and 2014. Another 302,236 agents appearing as notifiers were not registered.

Some values are logically or biologically impossible (i.e. gender=male AND pregnant=yes). The following most



**Figure 2.** Relation between birth date (DT\_NASCI) data filling and malaria cycle. This figure is in color in the electronic version.

frequently used variables were analyzed, relating pairs to check the coherence of values.

The follow-up consultation was related to the active/passive notification. The patient's follow-up consultation occurs weeks after the confirmation of the malaria infection (World Health Organization 2012). It is a second consultation to certify the healing and is registered as a new notification, which is why it is essential to distinguish it from the first consultation to avoid duplicates. Two variables - the detection type, called TIPO\_LAM (1-Passive, 2-Active and 3- Follow-up consultation) and the follow-up consultation, called ID\_LVC (1-Yes, 2-No) certify the follow-up consultation. There were 201 inconsistent notifications checked as a follow-up consultation notification (TIPO\_LAM=3) but also as its opposite (ID\_LVC=2).

The age writing format was related to the patient age. Both variables work together. The age writing format shows whether the patient age value is registered in days, months or years. There were 299 values showing patients aged between 101 and 1004.

The patient age was related to the pregnancy length. There were seven patients less than one year old showing as pregnant between the 1<sup>st</sup> and 3<sup>rd</sup> trimester of pregnancy, 118 notifications of 10-year old patients with ignored gestational age, and 38 pregnant but over 55 years old.

The exam, the notification and the symptoms dates were related. Between the malaria infection and the cure, the order of events was not always symptom, then notification, and finally exam. When comparing the dates of these events, some behaviors were unexpected. For not-followed-up consultation notifications (1,530,499 notifications) there were 348,292 patients notifying the same day they reported the malaria symptoms and 1,151,271 patients who delayed notifying it until one week later. On the other hand, 13,597 persons

notified before having symptoms, which is not logical and 6,290 persons notified 60 days after the symptoms.

There were 1,395,019 (91.15%) clinical exams carried out on the same day as the notification and 9,276 (0.6%) realized one week afterwards. However, there were 60 unexpected clinical exams performed before the notification.

All those cases were concentrated mostly in the municipalities of Manaus and Coari around 2003.

The SIVEP-Malaria georeferencing system is organized by localities and the municipalities are responsible for updating this information. One locality can accommodate a health unit, a patient's residence or his declared infection site. The localities module provides the georeferenced points. However, only 1.06% of all 525,795 of SIVEP-Malaria's localities were actually georeferenced, of which 28.6% were on the Amazonas state, the most covered one.

### Data cleaning product

In order to maximize the capacity of analysis of SIVEP-Malaria, a data cleaning was made covering Amazonas state from 2003 to 2014. The goal was to identify the fully filled notifications with georeferenced health units, residences and infection localities. The data cleaning started excluding the 10 variables having the larger amount of missing values (Table 2) and the three ones that were replaced (Group B). The data entry included 1,801,229 malaria positive notifications. At first, 49.8% of notifications were excluded for having missing values. Next, less than 0.01% were excluded for having outliers. Then, 0.02% were removed for being inconsistent. Finally, the georeferencing of localities was analyzed. There were 2.03% of notifications without georeferenced health units, 37.59% of notifications without georeferenced patient residences and 44.82% of notifications without georeferenced infection localities.

The data cleaning product is summarized in Figure 3 as an erroneous data concentration map by municipality. The missing data, outliers and inconsistencies were counted for each notification and the result was matched with its respective notification municipality geocode to quantify the erroneous data occurrence by municipality. The total of invalid data was divided by the total of notifications for all years and multiplied by 100 returning the so called “erroneous data index”. Values varied from 4.47 for the municipality of São Gabriel da Cachoeira to 153.93 for Novo Airão.

After the data clearance and the subsequent selection of notifications based on georeferenced localities, 49.17% of data entry was exploitable for a study based on the health units, 31.33% for a study based on the patient’s residence and 27.70% for one based on the infection site.

**Impact on research**

Using the Google Scholar, the Malaria Journal, and the Web of Science internet search engines, and giving “SIVEP” and “Malaria” as keywords, we have counted a total of 349 publications: 280 articles, 14 books and 18 master/PhDs from

2003 to 2014. The number of publications gradually grows every year. The most used variables from SIVEP-Malaria were those used to elaborate the Annual Parasite Incidence (API) employed by the National Plan to Malaria Control (PNCM) to set transmission risk and priority areas for disease control. It is a reference index, often used at the municipal scale, and composed of four SIVEP-Malaria variables: exam results, municipality of residence, notification date and follow-up consultation. Those were also on the top of the list of used variables when used separately. The next most used variables are the residence locality, the municipality of notification, gender, age, the infection locality, the pregnancy length, active/passive notification, the treatment date and the symptom date.

**DISCUSSION**

Medical datasets often have missing and erroneous values due to a bad record or human error (Bai *et al.* 2015). There was a major difference of data availability between the 43 variables that makes it impossible to compare all data for every patient. Missing data formed the first step of the data cleaning,

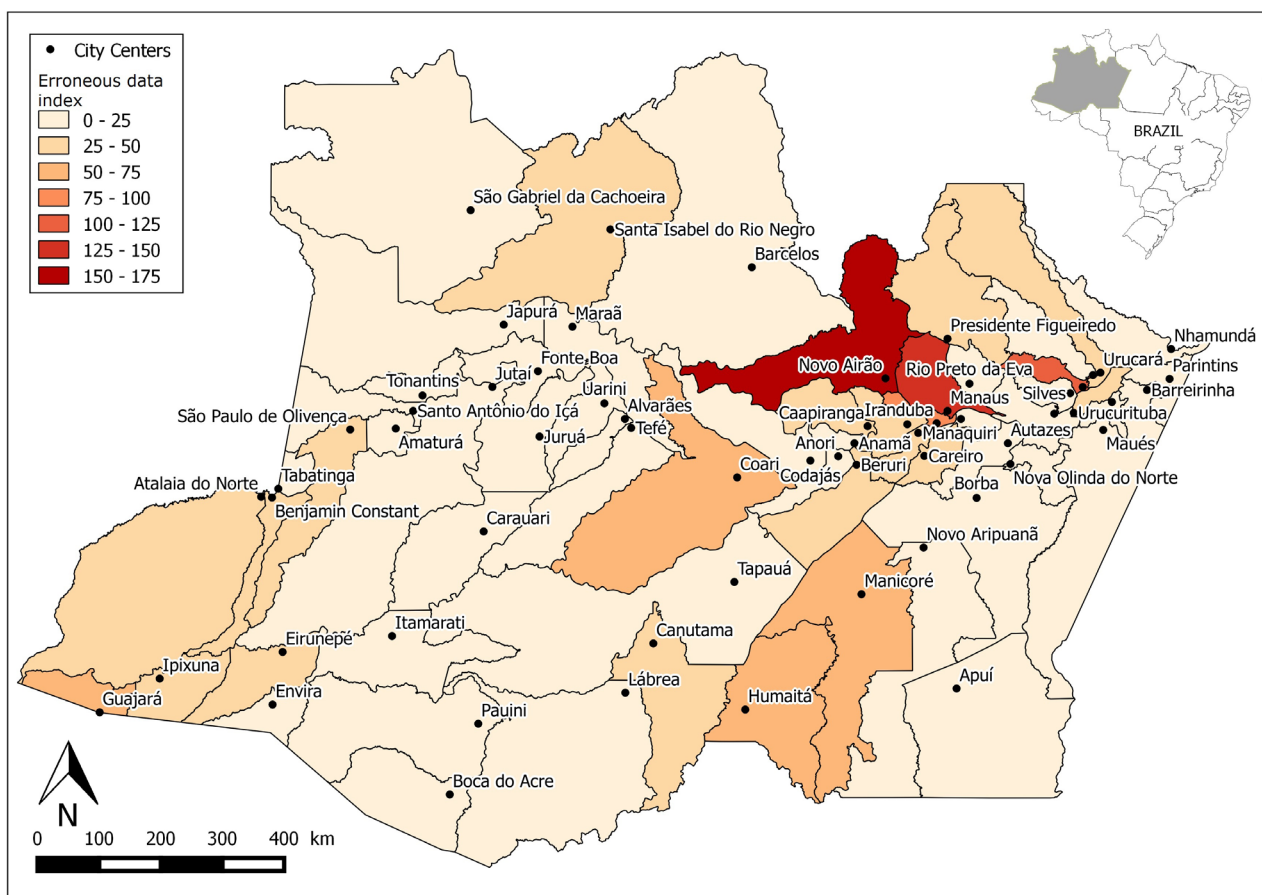


Figure 3. Erroneous data by Amazonas state municipalities from 2003 to 2014. This figure is in color in the electronic version.

discarding 49.8% of patients' notifications if the research demanded having all the variables. The variables from *Group A* are often responsible for this. Usually a personalized data cleaning has to be made for each study discarding undesirable variables before discarding missing data notifications. It is recommended to make a descriptive analysis and exploration to identify for each variable its availability for specific region and period.

Using the raw SIVEP-Malaria data rather than available processed products allows multiple approaches of the epidemiological reality and the method to be used has to consider the spatial and temporal scale to be analyzed. Local scale studies are often impossible, as for the municipalities of Itapiranga, Urucurituba, Nhamundá, Nova Olinda do Norte, Barreirinha, Boa Vista do Ramos and Silves, which accumulates zero georeferenced localities. The lack of fine geographical information explains why researches are made on a municipal scale. The georeferenced SIVEP localities points are missing for most parts of the territory and the provided population values are outdated. This disables a demographic dependent approach like the Annual Parasite Incidence (API) (Ray and Beljaev 1984) on a smaller scale than the municipal, forcing SIVEP-Malaria to be combined with the municipal demographic data of the Brazilian Institute of Geography and Statistics (IBGE).

Multi-temporal studies may contain errors if they exclude notifications with missing data from some erroneous variable, such as the new variables from *Group A*. This would lead to underestimation of the real amount of malaria cases. A study based on the ethnics is only possible after 2011. Using the data from *Group B* is also a mistake since they have been replaced, though they still have some insertions after 2011 that are possibly complementary to variables from *Group C*.

Compulsory reporting produces a great amount of data. SIVEP-Malaria gets data from 805 municipalities in nine states. Invalid data may occur, but we observe improvement in terms of data production and quality. It is possible to mark exactly where and when invalid data occurs and some are concentrated in certain municipalities or localities and also in specific time periods. Manaus and Coari are the two municipalities highlighted by having more invalid data. The earliest years of the system are clearly a transition phase to an operational system. The variables that are more constant in providing technically correct data are the variables of exclusive responsibility of the medical unit/agent. In second place is the information provided by the patient.

SIVEP-Malaria is in continuous development. It will only be fully functional when all variables are systematically filled and when the complementary modules, like the list of georeferenced locations, health agents or examiners, are updated by municipalities. The outliers and the

inconsistencies are not very important statistically when looking at the whole region of the Amazonas state, but can be a serious threat to a local study for specific periods, since the accumulation of erroneous data can be precise in space and time. Erroneous data may come from the patient declaration, the health agent reporting or from mistyping while uploading to the database. Visualizing the data behavior through space and time allows one to learn much about the database and the collecting system, and enables one to adjust the methodology of research to avoid unexpected situations. SIVEP-Malaria still lacks strategic data, such as geographical, and the fact that it has passed through major changes in its young lifetime does not allow its full temporal capacity to be used for some variables. However, it has a great capacity to provide information on the distribution and trends in malaria, especially at the municipal scale or using the georeferenced medical unit for the local scale. Although, the system is designed so as to, eventually give a good data quality to the residential and the infection site localities. It is already a great data source for scientific research, even though it has been created for monitoring and administration.

## CONCLUSIONS

The results of the SIVEP-Malaria data base data cleaning for the state of Amazonas show that some variables are best to be used than others. The major limitation in using variables is the absence of data in certain places, and certain periods, which could affect comparisons between places and dates. Plotting data availability by variable over the years allowed a global view of the most robust SIVEP-Malaria variables. We detected several errors, as unrealistic outliers, and incoherent data. The source of the detected erroneous data can be identified, and most part originates from the municipalities of Coari and Manaus, mainly from the early dates of the SIVEP-Malaria. The errors can also be influenced by the epidemic fluctuations. The number of studies that uses SIVEP data increases every year. A best knowledge of the structure of this data base is critical. The results shows that it is highly recommend to realize the applied pretreatment method before performing statistics with the data.

## ACKNOWLEDGEMENT

This research was partially supported by the Ciência sem Fronteiras (CsF), Brazilian Scholarship Program, the National Institute of Science and Technology Innovation for Neglected Diseases (INCT-IDN) and the Observatory of Climate and Health from Oswaldo Cruz Foundation in Rio de Janeiro, Brazil and the Unité Mixte de Recherche Espace pour le Développement (UMR ESPACE-DEV) from Montpellier, France.

## REFERENCES

- Almeida, L.B.; Barbosa M.G.V.; Espinosa, F.E.M. 2010. Malária em mulheres de idade de 10 a 49 anos, segundo o SIVEP- Malária, Manaus, Amazonas, 2003-2006. *Revista da Sociedade Brasileira de Medicina Tropical*, 43: 304-308.
- Bai, E.W.; Johnson, H.; Xu, W.; Jacob, M. 2015. A preliminary study on cleaning up erroneous data and filling in missing values in a medical record. *IFAC-PapersOnLine*, 48(20): 493-498.
- Barata, R.C.B. 1995. Malaria in Brazil: Trends in the last ten years. *Cadernos de Saúde Pública*, 11: 128-136.
- Brasil, 2003. Relatório de monitoramento. Programa Nacional de Controle da Malária. *Tribunal de Contas da União*, Brasília, 27p.
- Brasil, 2014. Portaria nº 1.271, de 6 de junho de 2014. Brasília.
- Brasil, 2015. Departamento de informática do Sistema Único de Saúde do Brasil. (<http://www2.datasus.gov.br/DATASUS/>). Accessed on 09/02/2015.
- Braz, R.; Andreozzi, V.; Kale P. 2006. Detecção precoce de epidemias de malária no Brasil: Uma proposta de automação. *Epidemiologia e Serviços de Saúde*, 15: 21-33.
- Broeck, J. van der.; Cunningham, S.A.; Eeckels, R.; Herbst, K. 2005. Data cleaning: detecting, diagnosing, and editing data abnormalities. *PLoS Medicine* 2: e267.
- Costa, K.M.M.; Almeida W.A.F.; Magalhães, I.B.; Montoya, R.; Moura, M.S.; Lacerda, M.V.G. 2010. Malária em Cruzeiro do Sul (Amazônia Ocidental brasileira): Análise da série histórica de 1998 a 2008. *Revista Panamericana de Salud Pública*, 28: 353-60.
- Deane, L. M. 1988. Malaria studies and control in Brazil. *The American journal of tropical medicine and hygiene*, 38: 223-30.
- Ferreira, M. U.; Silva-Nunes, M. 2010. Evidence-based public health and prospects for malaria control in Brazil. *Journal of Infection in Developing Countries*, 4: 533-545.
- Hang-Hai, D.; Erhard, R. 2000. Data Cleaning: Problems & Current Approaches. *IEEE bulletin of the technical committee on Data Engineering*, 23.
- Hermes, S.C.N.M.; Nunes, V.L.B.; Dorval, M.E.C.; Brillhante, A.F. 2013. Aspectos epidemiológicos da malária humana no município de aripuanã, estado de Mato Grosso, Brasil, 2005 a 2010. *Hygeia*, 9: 42-51.
- Jonge, E.D.; van der Loo, M. 2013. An introduction to data cleaning with R. Statistics Netherlands, *The Hague*, 53p.
- Martens, P.; Kovats, R.S.; Nijhof, S.; Vries, P.; Livermore, M.T.J.; Bradley, D.J.; Cox, J.; McMichael, A.J. 1999. Climate change and future populations at risk of malaria. *Global Environmental Change*, 9: 89-107.
- Oliveira-Ferreira, J.; Lacerda, M.V.G.; Brasil, P.; Ladislau, J.L.B.; Tauil, P.L.; Ribeiro, C.T.D. 2010. Malaria in Brazil: An overview. *Malaria Journal*, 9: 115
- Peiter, P.C.; Franco, V.C.; Gracie, R.; Xavier, D.R.; Mutis, M.C.S. 2013. Situação da malária na tríplice fronteira entre Brasil, Colômbia e Peru. *Cadernos de Saúde Pública*, 29: 2497-2512.
- Pina-Costa, A.; Brasil, P.; Santi, S.; Araujo, M. P.; Suárez-Mutis, M. C.; Santelli, A. C. F. S. 2014. Malaria in Brazil: what happens outside the Amazonian endemic region. *Memórias do Instituto Oswaldo Cruz*, 109: 618-633.
- Ray, A. P.; Beljaev, A. E. 1984. Epidemiological surveillance: a tool for assessment of malaria and its control. *The Journal of communicable diseases*, 16: 197-207.
- Sampaio, V.S.; Siqueira, A.M.; Alecrim, M.G.C.; Mourão, M.P.G.; Marchesini, P.B.; Albuquerque, B.C.; et al. 2015. Malaria in the State of Amazonas: A typical Brazilian tropical disease influenced by waves of economic development. *Revista da Sociedade Brasileira de Medicina Tropical*, 48: 4-11.
- Schoenbach, V.J.; Rosamond, W.D (eds). 2000. Understanding the fundamentals of epidemiology: An evolving text. Chapel Hill, NC: University of North Carolina at Chapel Hill. *Fall 2000 Edition*, 481p.
- Silva, N. S.; Silva-Nunes, M.; Malafronte, R.S.; Menezes, M.J.; D'Arcadia, R.R.; Komatsu, N.T.; Scopel K. K. G.; Braga, E. M.; Cavasini, C. E.; Cordeiro, J. A.; Ferreira, M.U. 2010. Epidemiology and control of frontier malaria in Brazil: lessons from community-based studies in rural Amazonia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 104: 343-50.
- Tauil, P.L.; Daniel-Ribeiro, C. 1998. Some aspects of epidemiology and control of malaria in Brazil. *Research and Reviews in Parasitology*, 58:163-167.
- World Health Organization. 2012. Disease surveillance for malaria control. An operational manual. *WHO Press*, Geneva, 84p.

Recebido em 08/02/2016

Aceito em 13/05/2016