

**UNIVERSIDADE DE BRASÍLIA**  
**FACULDADE DE MEDICINA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM MEDICINA TROPICAL**

**KLAUSS KLEYDMANN SABINO GARCIA**

**O progresso do Brasil na eliminação da malária: análise de séries  
temporais de casos de malária e estudo de caso-controle para fatores  
associados ao óbito por malária no Brasil de 2011 a 2022**

**BRASÍLIA**

**2023**

**O progresso do Brasil na eliminação da malária: análise de séries temporais de casos de malária e estudo de caso-controle para de fatores associados ao óbito por malária no Brasil de 2011 a 2022**

**Klauss Kleydmann Sabino Garcia**

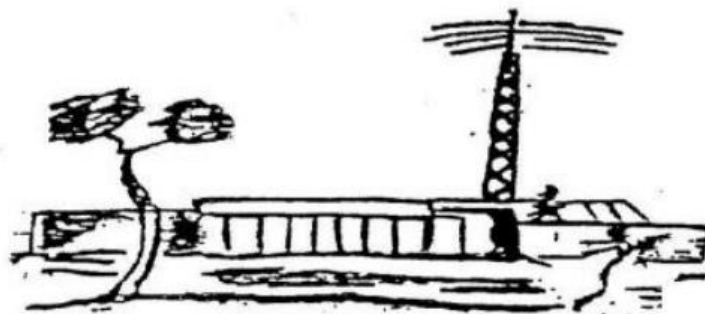
Tese de doutorado apresentada ao Programa de Pós-Graduação em Medicina Tropical da Universidade de Brasília, como requisito para obtenção do título de Doutor em Medicina Tropical.

Área de concentração: Epidemiologia e Controle de Doenças Infecciosas e Parasitárias

Linha de pesquisa: Estudo do adoecimento e morte por doenças infecciosas e parasitárias

**Orientador:** Walter Massa Ramalho

**Co-orientador:** André Machado de Siqueira



**BRASÍLIA**

**2023**

Ficha catalográfica elaborada automaticamente,  
com dados fornecidos pelo autor.

Garcia, Klauss Kleydmann Sabino

O progresso do Brasil na eliminação da malária: análise de séries temporais de casos de malária e estudo de caso-controle para fatores associados ao óbito por malária no Brasil de 2011 a 2022./ Klauss Kleydmann Sabino Garcia; orientador Walter Massa Ramalho - - Brasília, 2023.

200 p.

Tese (Doutorado - Mestrado em Medicina Tropical) -- Universidade de Brasília, 2023.

1. Malária. 2. Saúde Pública. 3. Epidemiologia. 4. Análises de séries temporais. 5. Estudos Caso-controle. I. Ramalho, Walter Massa, orient. II. Siqueira, André Machado, co-orient. III. Título.

**O progresso do Brasil na eliminação da malária: análise de séries temporais de casos de malária e estudo de caso-controle para fatores associados ao óbito por malária no Brasil de 2011 a 2022**

**DATA DA DEFESA**

**06 de novembro de 2023**

**COMPOSIÇÃO DA BANCA DE QUALIFICAÇÃO**

**Presidente da banca**

Doutor Walter Massa Ramalho  
Universidade de Brasília

**Membro interno**

Doutora Elisabeth Duarte  
Universidade de Brasília

**Membro externo**

Doutor Marcus Vinícius Guimarães de Lacerda  
Fundação de Medicina Tropical Doutor Heitor Vieira Dourado

**Membro externo**

Doutor Gabriel Zorello Laporta  
Centro Universitário Faculdade de Medicina do ABC

**Membro suplente**

Doutora Sheila Rodrigues Rodovalho  
Organização Pan-Americana da Saúde (OPAS)

## **AGRADECIMENTOS**

A Deus, por ter me proporcionado oportunidades que me permitiram chegar até aqui, e continuar.

À minha esposa – Amanda Amaral Abrahão, por ser minha companheira de vida, de serviço e de pesquisa, por me incentivar e torcer pelo nosso crescimento.

A meus pais, por terem me criado com a frase “Nunca pare de estudar”, e hoje estudar, para mim, é um estilo de vida.

Ao meu orientador Walter Massa Ramalho pela orientação e liberdade no doutorado, que me permitiu aprender a ser um pesquisador.

Ao meu orientador André Machado de Siqueira, que me abriu portas que eu jamais imaginara que existiam, por me incentivar a pensar maior e a provocar mudanças no mundo.

Aos queridos membros da banca: Professor Marcus Lacerda, Professora Beth Duarte, Professor Laporta e Doutora Sheila Rodrigues Rodvalho. Os quais tenho profunda admiração e carinho. Agradeço por terem aceitado participar deste momento e por, durante minha formação, terem sido – além de pesquisadores – professores.

Também às instituições que tornaram possível a conclusão deste trabalho: London School of Hygiene & Tropical Medicine (LSHTM), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Universidade de Brasília - Medicina Tropical (UnB - Medicina Tropical).

*Todo o bem que fazemos pelos outros  
um dia será feito por nós – Kamado Tanjiro*

## LISTA DE FIGURAS

### *Referencial Teórico e Métodos*

Figura 1. Países do mundo com presença de casos autóctones de malária e sua situação em 2021.....	21
Figura 2. Incidência (1.000 hab.) e coeficiente de mortalidade de malária (100 mil hab.) no mundo de 2000 a 2021.....	22
Figura 3. Incidência (1.000 hab.) e coeficiente de mortalidade de malária (100 mil hab.) na região das Américas de 2000 a 2021. ....	23
Figura 4. Séries histórica de casos de malária no Brasil, 1959 a 2020. ....	25
Figura 5. Incidência Parasitária Anual de malária no Brasil, 2021. ....	27

### *Resultados - Objetivo 1*

Figura 6. Taxa de casos, óbitos e letalidade por malária na região extra- Amazônica do Brasil, 2011 a 2020. ....	56
Figura 7. Fluxo de casos importados de malária notificados na região extra-Amazônica do Brasil com taxas de incidência. 2011 a 2020. ....	57
Figura 8. Atributos de série temporal, tendência, sazonalidade e aleatoriedade de casos notificados de malária, por mês, na região extra-Amazônica brasileira com previsão de dois anos. ....	59

### *Resultados - Objetivo 2*

Figura 9. Distribuição Espacial de Todas as Classificações de Risco de Infecções por Malária na Amazônia Brasileira, 2021-2022. ....	76
Figura 10. Distribuição Espacial de Classificações de Risco de Infecções por Plasmodium falciparum na Amazônia Brasileira, 2021-2022. ....	77
Figura 11. Tendências de casos de malária na Amazônia Brasileira, 2011-2022... 79	
Figura 12. Todos os casos autóctones de malária na Amazônia Brasileira, previsão para 2023 e 2024 .....	80

Figura 13. Casos autóctones de malária nos estados endêmicos da Amazônia Brasileira, previsão para 2023.....	82
--	----

*Resultados - Objetivo 3*

Figura 14. Fluxograma do processo de linkage probabilístico .....	100
---	-----

*Resultados - Objetivo 4*

Figura 15. Definição de casos e controles de acordo com o processo de vinculação de dados. ....	121
Figura 16. Óbitos por malária no Brasil de 2011 a 2020.....	124

## LISTA DE TABELAS

### *Referencial Teórico e Métodos*

Tabela 1. Perfil epidemiológico dos casos de malária na região extra-AMZ do Brasil, 2011 a 2020. ....	53
---	----

### *Resultados - Objetivo 1*

Tabela 2. Distribuição dos casos por espécie de Plasmodium, atividade profissional e autoctonia, 2011 a 2020, região extra-AMZ do Brasil. ....	55
Tabela 3. Quantidade de pareamentos por chaves-de-vinculação utilizadas.....	101

### *Resultados - Objetivo 3*

Tabela 4. Perfil da mortalidade por malária e dos óbitos pareados entre 2011 e 2020, Brasil.....	102
Tabela 5. Causas básicas de óbito (SIM) e recuperação de resultados laboratoriais (Sivep-Malária e Sinan).....	104
Tabela 6. Perfil Epidemiológico dos óbitos por malária no Brasil, 2011-2020.....	127

### *Resultados - Objetivo 4*

Tabela 7. Regressão logística múltipla de óbitos por malária na região Amazônica brasileira, 2011-2020.....	129
Tabela 8. Regressão logística múltipla de óbitos por malária na região extra-Amazônica brasileira, 2011-2020. ....	130

## **ÓRGÃOS FINANCIADORES**

Parte deste trabalho foi financiado pela Fundação Bill e Melinda Gates [INV-003970], pelo Ministério da Saúde/DECIT/CNPq (#443148/2019-8) e pela Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES – 10/2022 #88881.690401/2022-01).

## LISTA DE ABREVIATURAS E SIGLAS

AMZ – Região Amazônica  
APC – Annual Percent Change  
CAPES – Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior  
CID – Classificação Internacional de Doenças  
CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico  
CPF – Cadastro de Pessoas Físicas  
DATASUS – Departamento de Informática do Sistema Único de Saúde  
ETG – Estratégia Técnica Global  
EWMA – Média móvel ponderada exponencialmente  
Extra-AMZ – Região extra-Amazônica  
Fiocruz – Fundação Oswaldo Cruz  
FIOTEC – Fundação de apoio à Fiocruz  
HW – Holt-Winters  
IBGE – Instituto Brasileiro de Geografia e Estatística  
IC – Intervalo de Confiança  
IPA – Incidência Parasitária Anual  
LGPD – Lei Geral de Proteção de Dados  
LPI – Local de provável infecção  
LVC – Lâmina de verificação de cura  
MILD – Mosquiteiro Impregnado com Inseticida de Longa Duração  
MILDS – Mosquiteiros Impregnados com Inseticida de Longa Duração  
MPC – Monthly Percent Change  
MS – Ministério da Saúde  
ODS – Objetivos do Desenvolvimento Sustentável  
OMS – Organização Mundial da Saúde  
ONU – Organização das Nações Unidas  
OPAS – Organização Pan-Americana da Saúde

OR – Odds-Ratio

*P. falciparum* – *Plasmodium falciparum*

*P. malariae* – *Plasmodium malariae*

*P. vivax* – *Plasmodium vivax*

*Pf* – *Plasmodium falciparum*

*Pm* – *Plasmodium malariae*

PNCM – Programa Nacional de Prevenção e Controle da Malária

*Po* – *Plasmodium ovale*

PQA-VS – Programa de Qualificação das Ações de Vigilância em Saúde

*Pv* – *Plasmodium vivax*

RG – Registro Geral (identidade)

SIM – Sistema de Informação de Mortalidade

Sinan – Sistema de informações de agravos de notificação

SIS – Sistemas de Informação em Saúde

SISMAL – Sistema de Informações do Programa Nacional de Controle Integrado da Malária

Sivep-Malária – Sistema de Informação de Vigilância Epidemiológica da Malária

SUS – Sistema Único de Saúde

SVSA – Secretaria de Vigilância em Saúde e Ambiente

TDR – Testes de diagnóstico rápido

UF – Unidades Federativas

## SUMÁRIO

<b>RESUMO</b> .....	xv
<b>ABSTRACT</b> .....	xvii
<b>INTRODUÇÃO</b> .....	19
<b>1. REVISÃO DA LITERATURA</b> .....	21
1.1 Epidemiologia da malária no mundo e no Brasil.....	21
1.2 Diagnóstico.....	27
1.3 Tratamento .....	28
1.4 A Vigilância epidemiológica da malária .....	29
1.4.1 Objetivos e características da Vigilância epidemiológica.....	29
1.4.2 Sistemas de Informação em Saúde para malária no Brasil.....	30
1.4.3 Estratégias de controle da malária .....	32
1.5 O plano de eliminação da malária no Brasil .....	33
1.6 Vinculação de bancos de dados.....	35
1.7 Modelos de predição temporal .....	37
<b>2. JUSTIFICATIVA</b> .....	38
<b>3. OBJETIVOS</b> .....	39
3.1. Objetivo geral .....	39
3.2. Objetivos específicos.....	39
<b>4. MATERIAIS E MÉTODOS</b> .....	40
4.1. Aspectos metodológicos gerais .....	40
4.2. Aspectos metodológicos específicos .....	40
4.2.1. Descrição cenário epidemiológico da malária na região extra- Amazônica brasileira (Objetivo específico 1) .....	40
4.2.2. Descrição do cenário epidemiológico da malária na região Amazônica brasileira (Objetivo específico 2) .....	42
4.2.3. Descrição dos aspectos metodológicos de <i>record linkage</i> para sistemas de informação de malária (objetivo 3) .....	45
4.2.4. Descrição dos fatores associados ao óbito por malária (objetivo 4)	47
<b>5. RESULTADOS</b> .....	49
5.1. Descrição do comportamento e cenário epidemiológico da malária na região extra-Amazônica brasileira (Objetivo específico 1) .....	49

5.1.1. Série temporal da malária extra-AMZ no Brasil: cenário epidemiológico e um modelo de previsão de dois anos. (*Artigo publicado no Malaria Journal - Apêndice II) .....	49
5.2. Descrição do comportamento e cenário epidemiológico da malária na região Amazônica brasileira (Objetivo específico 2) .....	69
5.2.1. O Brasil está no caminho de eliminar a malária? Uma análise de séries temporais de casos de malária de 2011 e 2022.....	69
5.3. Descrição dos aspectos metodológicos de <i>record linkage</i> para sistemas de informação em saúde de malária (Objetivo específico 3).....	92
5.3.1. O uso de <i>record linkage</i> para recuperação de dados para vigilância de óbitos por malária.....	92
5.4. Descrição dos fatores associados ao óbito por malária (Objetivo específico 4).....	116
5.4.1. Fatores associados à morte por malária no Brasil: um estudo de caso-controle de 2011 a 2020.....	116
<b>6. CONSIDERAÇÕES FINAIS .....</b>	<b>143</b>
<b>7. REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>146</b>
<b>8. ANEXO I .....</b>	<b>156</b>
<b>9. ANEXO II .....</b>	<b>161</b>
<b>10. ANEXO III .....</b>	<b>162</b>
<b>11. ANEXO IV .....</b>	<b>163</b>
<b>12. ANEXO V .....</b>	<b>174</b>
<b>13. ANEXO VI .....</b>	<b>187</b>

## RESUMO

**INTRODUÇÃO:** A malária é uma doença infecciosa aguda predominante em países tropicais e em desenvolvimento. De acordo com a Organização Mundial da Saúde (OMS), em 2021, ocorreram mais de 247 milhões de casos e mais de 627 mil óbitos. No Brasil, no mesmo ano, foram registrados 137 mil casos e 61 óbitos. A região amazônica brasileira contabiliza 99% dos casos autóctones do país; no entanto, a região extra-Amazônica apresenta maior letalidade pela doença. Portanto, considerando a vigilância epidemiológica brasileira e o objetivo de eliminar a malária, conforme definido pela Estratégia Técnica Global e pelos Objetivos do Desenvolvimento Sustentável, o Brasil visa eliminar a transmissão autóctone do país até 2035 e reduzir os óbitos até 2030. Diante disso, este trabalho tem como objetivo descrever e analisar o cenário epidemiológico da malária no Brasil no período de 2011 a 2022, identificando os fatores associados à ocorrência de óbitos por malária no país.

**OBJETIVOS:** Descrever e analisar o cenário epidemiológico da malária no Brasil de 2011 a 2022 e identificar os fatores associados aos óbitos por malária no Brasil. **MÉTODOS:** São quatro estudos que analisam dados de casos e óbitos de malária entre 2011 e 2022. Dois desses estudos possuem um desenho transversal, que descreve e caracteriza os perfis e cenários epidemiológicos da região AMZ e da região extra-AMZ brasileira, incluindo o desenvolvimento de modelos preditivos (forecast). O terceiro estudo, de natureza transversal e caráter metodológico, relata os processos de vinculação de dados (*record linkage*) que podem contribuir para a qualificação das informações de óbitos por malária e melhorar as atividades de investigação de óbitos. O quarto estudo é do tipo caso-controle, no qual se descreve o perfil epidemiológico dos óbitos por malária entre residentes do Brasil e se investigam os fatores associados a essas mortes. Para os dois primeiros estudos, foram utilizados dados abertos, enquanto para os dois últimos, foram utilizados dados nominais. A pesquisa foi aprovada pelo Comitê de Ética em Pesquisa da Faculdade de Medicina da Universidade de Brasília.

**RESULTADOS:** Na região extra-AMZ, identificou-se um predomínio de casos importados de outros países e da própria região AMZ brasileira, principalmente relacionados às atividades laborais dos indivíduos. Na região AMZ, predominam os casos de transmissão local, e o cenário prospectivo sugere a possibilidade de eliminação da malária no Brasil até 2035. No entanto, isso requer superar desafios, especialmente nos estados do Amazonas e Roraima, bem como nas áreas de garimpo e entre as populações indígenas. Os óbitos no país têm aumentado desde 2019, com uma letalidade maior na região extra-AMZ brasileira, decorrente principalmente de casos de malária importada de outros países e de infecções com alta parasitemia, causadas por espécies como *P. falciparum*, *P. ovale* e *P. malariae*. Na região AMZ, os óbitos são predominantemente devido a infecções por *P. vivax*, com fatores de risco associados a atrasos na detecção/diagnóstico, infecções em áreas indígenas, infecções em crianças menores de um ano e em adultos com mais de 40 anos, alta parasitemia e detecção passiva do caso.

**CONCLUSÕES:** Esta tese proporcionou uma compreensão abrangente do cenário e do perfil epidemiológico da malária no Brasil, destacando as variações temporais ao longo dos anos e as tendências futuras de ocorrência de casos. Além disso, as técnicas de vinculação de dados permitiram investigar causas relacionadas e associadas aos óbitos por malária, oferecendo perspectivas sobre como esse conhecimento e essas técnicas podem fortalecer as políticas públicas de saúde e auxiliar o Brasil na sua meta de eliminar a malária até 2035.

**Palavras-chave:** Malária; Vigilância epidemiológica; Análises de Séries temporais; Saúde Pública; Epidemiologia.

## ABSTRACT

**INTRODUCTION:** Malaria is an acute infectious disease prevalent in tropical and developing countries. According to the World Health Organization (WHO), in 2021, there were over 247 million cases and more than 627,000 deaths globally. In Brazil, during the same year, 137,000 cases and 61 deaths were recorded. The Brazilian Amazon region accounts for 99% of the country's autochthonous cases; however, the extra-Amazon region shows higher lethality from the disease. Therefore, considering Brazilian epidemiological surveillance and the goal of malaria elimination set by the Global Technical Strategy and the Sustainable Development Goals, Brazil aims to eliminate autochthonous transmission by 2035 and reduce deaths by 2030. Considering this, this study aims to describe and analyze the epidemiological scenario of malaria in Brazil from 2011 to 2022, identifying factors associated with malaria-related deaths in the country. **OBJECTIVES:** To provide a descriptive and analytical overview of the epidemiological landscape of malaria in Brazil from 2011 to 2022 and to identify factors associated with malaria-related fatalities in Brazil. **METHODS:** There are four studies analyzing malaria case and death data from 2011 to 2022. Two of these studies have a cross-sectional design, describing and characterizing the epidemiological profiles and scenarios of the AMZ and non- AMZ regions of Brazil, including the development of predictive models (forecasting). The third study is of cross-sectional and methodological nature, reports data *record linkage* processes that can contribute to improving the quality of malaria death information and enhance death investigation activities. The fourth study is a case-control study, describing the epidemiological profile of malaria deaths among Brazilian residents and investigating factors associated with these deaths. Open data were used for the first two studies, while nominal data were used for the latter two. The research was approved by the Research Ethics Committee of the School of Medicine at the University of Brasília. **RESULTS:** In the extra-AMZ region, a predominance of imported cases from other countries and the Brazilian AMZ

region was identified, mainly related to individuals' occupational activities. In the AMZ region, locally transmitted cases predominate, and the prospective scenario suggests the possibility of malaria elimination in Brazil by 2035. However, this requires overcoming challenges, especially in the states of Amazonas and Roraima, as well as in mining areas and among indigenous populations. Deaths in the country have been increasing since 2019, with higher lethality in the extra-AMZ region, mainly due to cases imported from other countries and high parasitemia infections caused by species such as *P. falciparum*, *P. ovale*, and *P. malariae*. In the AMZ region, deaths are predominantly due to *P. vivax* infections, with risk factors associated with delays in detection/diagnosis, infections in indigenous areas, infections in children under one year old and adults over 40 years old, high parasitemia, and passive case detection. **CONCLUSIONS:** This thesis provided a comprehensive understanding of the malaria scenario and epidemiological profile in Brazil, highlighting temporal variations over the years and future case occurrence trends. Furthermore, *record linkage* techniques allowed investigation of related and associated causes of malaria deaths, offering insights into how this knowledge and these techniques can strengthen public health policies and assist Brazil in its goal to eliminate malaria by 2035.

**Keywords:** Malaria; Epidemiological Surveillance; Time Series Analysis; Public Health; Epidemiology.

## INTRODUÇÃO

A malária é uma doença infecciosa febril aguda que predominantemente ocorre em países tropicais em desenvolvimento. De acordo com a Organização Mundial da Saúde (OMS), em 2021, afetou aproximadamente 247 milhões de pessoas em todo o mundo e resultou em mais de 627 mil óbitos. Apesar dos esforços globais para controlar a malária e suas conexões com os Objetivos do Desenvolvimento Sustentável (ODS) das Nações Unidas (ONU), a doença mantém taxas de incidência elevadas, com 59 casos a cada 1.000 habitantes, bem como taxas de mortalidade significativas, com 14,8 óbitos a cada 100.000 habitantes em todo o mundo (TALAPKO et al., 2019; OMS, 2022).

No Brasil, a malária representa um problema de saúde pública, principalmente na região da Amazônia (AMZ), onde cerca de 99% dos casos são registrados, em contraste com a região extra-Amazônica (extra-AMZ), que responde por menos de 1% dos casos novos de malária. Em 2021, segundo o Programa Nacional de Prevenção e Controle da Malária (PNCM) do Ministério da Saúde do Brasil (MS), foram notificados aproximadamente 137 mil casos novos e autóctones de malária (BRASIL, 2022A). O país apresentou uma forte redução no número de casos desde 2005, quando foram registrados cerca de 600 mil casos de malária. Essa redução foi contínua até 2017 e 2018, quando o Brasil vivenciou um aumento repentino no número de casos, com mais de 185 mil casos autóctones registrados (BRASIL, 2021A).

Em 2021 foram registrados 61 óbitos (BRASIL, 2023). Embora a quantidade de óbitos pela doença pareça baixa em comparação com a quantidade de casos, esses óbitos destacam as fragilidades do Sistema Único de Saúde (SUS) brasileiro, uma vez que as mortes por malária são consideradas evitáveis (MALTA e DUARTE, 2007). Na região da Amazônia Legal, a taxa de letalidade por malária é de 0,02%, enquanto na região extra-AMZ essa taxa pode chegar a 3,3%, ou seja, 165 vezes maior do que na região da Amazônia (BRASIL, 2021A).

O complexo cenário da malária no território nacional possui grandes lacunas dos conhecimentos quanto a causas relacionadas à ocorrência de óbitos. Estudos realizados em nível mundial indicam que óbitos por malária estão associados a fatores como sexo, condições pré-existentes, comorbidades, coinfeção por HIV/Aids, demora na busca por atendimento, tratamento não oportuno, tratamento inadequado, viagens para áreas endêmicas, sazonalidade e infecção por *P. falciparum* (DERESSA e FANTAHUN, 2007; SARKAR et al., 2009; LÜTHI e SCHLAGENHAUF, 2015).

Portanto, considerando a vigilância epidemiológica brasileira e a disponibilidade de testagem e tratamentos gratuitos, é preocupante que os óbitos por malária ainda ocorram no país. Diante disso, este trabalho tem como objetivo descrever e analisar o cenário epidemiológico da malária no Brasil no período de 2011 a 2022, identificando os fatores associados à ocorrência de óbitos por malária no país.

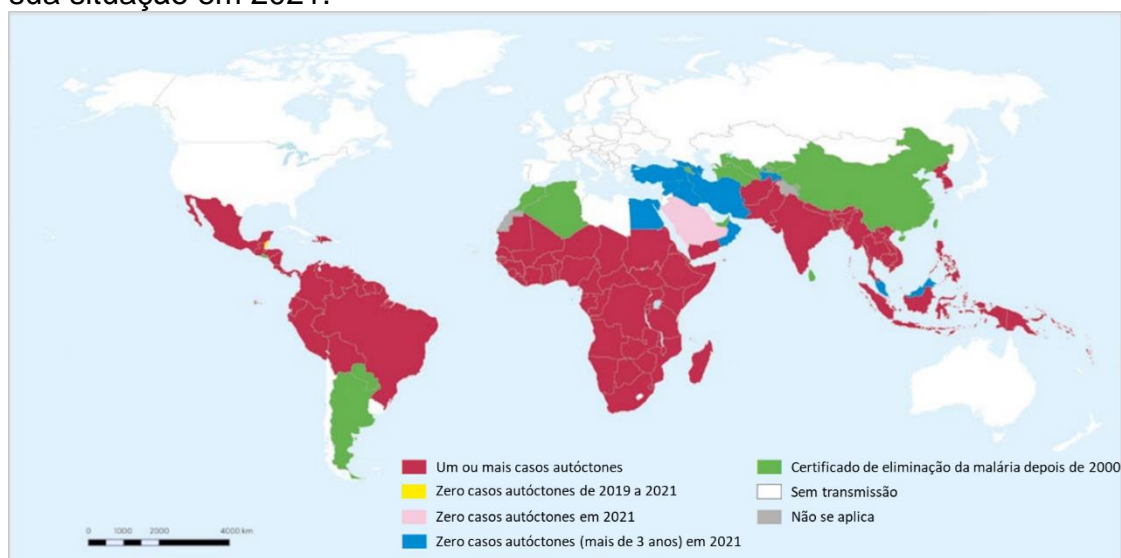
## 1. REVISÃO DA LITERATURA

### 1.1 Epidemiologia da malária no mundo e no Brasil

#### *No mundo e nas Américas*

A malária é uma doença que tem registros antigos, cujas manifestações clínicas foram descritas em épocas que antecedem o nascimento de Cristo (COURA et al., 2006). De acordo com a Organização Mundial da Saúde (OMS), em 2021 foram registrados mais de 247 milhões de casos de malária, em 84 países diferentes, com mais de 619 mil óbitos. Sua presença é predominante em países tropicais e a maior quantidade de países com transmissão ativa se concentram nas regiões da África, Ásia e Américas (América central e América latina) (figura 1).

Figura 1. Países do mundo com presença de casos autóctones de malária e sua situação em 2021.

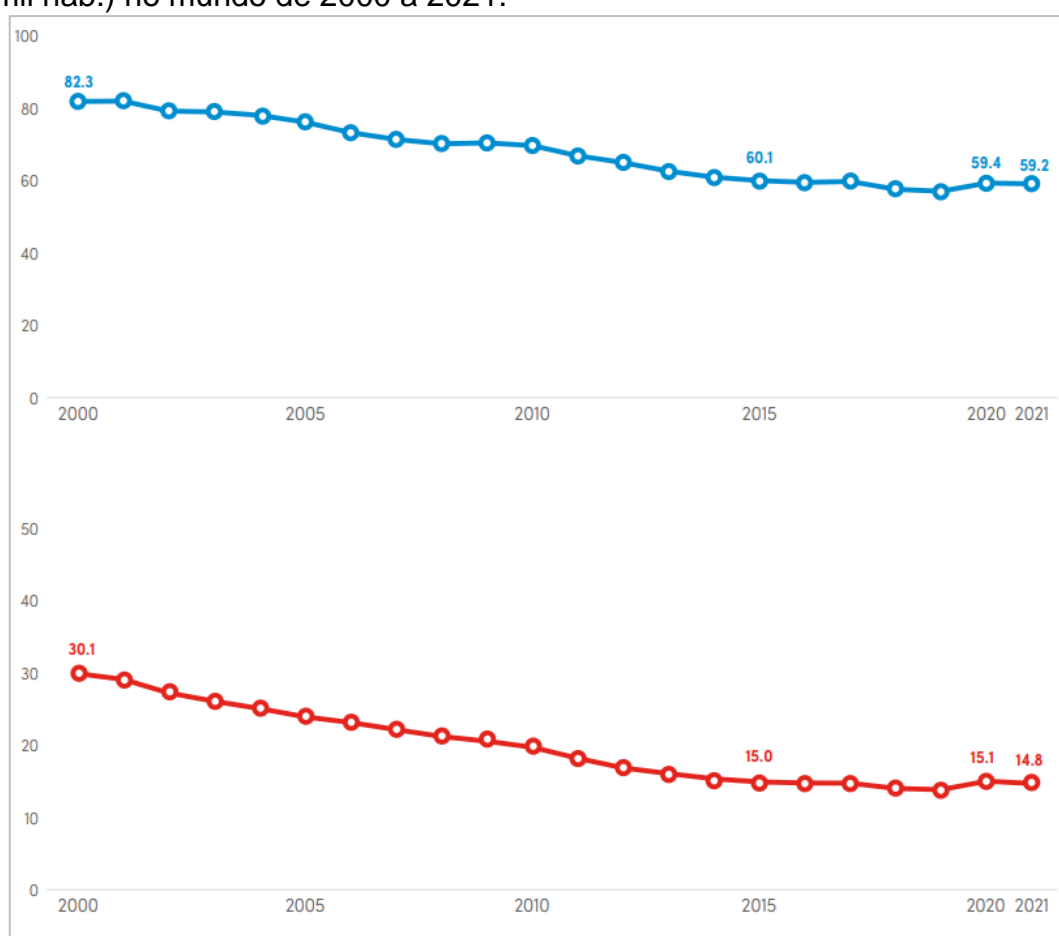


Fonte: World Malaria Report - OMS, 2022.

O total de casos em 2021 apresentou um aumento de 0,8% (2 milhões) no número de casos em relação a 2020. O total de óbitos apresentou uma redução de 0,96% (6 mil) em relação a 2020, com um coeficiente de

mortalidade de 14,8 óbitos para cada 100 mil habitantes, quando em 2020 era de 15 óbitos para cada 100 mil habitantes (OMS, 2022). O mundo estava apresentando reduções constantes na incidência parasitária anual e nos coeficientes de mortalidade até o ano de 2019. Em 2020 estes indicadores apresentaram aumento e isto pode estar relacionado a problemas estruturais dos serviços de saúde devido à pandemia de COVID-19 (OMS, 2021A; ULRICH et al., 2021).

Figura 2. Incidência (1.000 hab.) e coeficiente de mortalidade de malária (100 mil hab.) no mundo de 2000 a 2021.



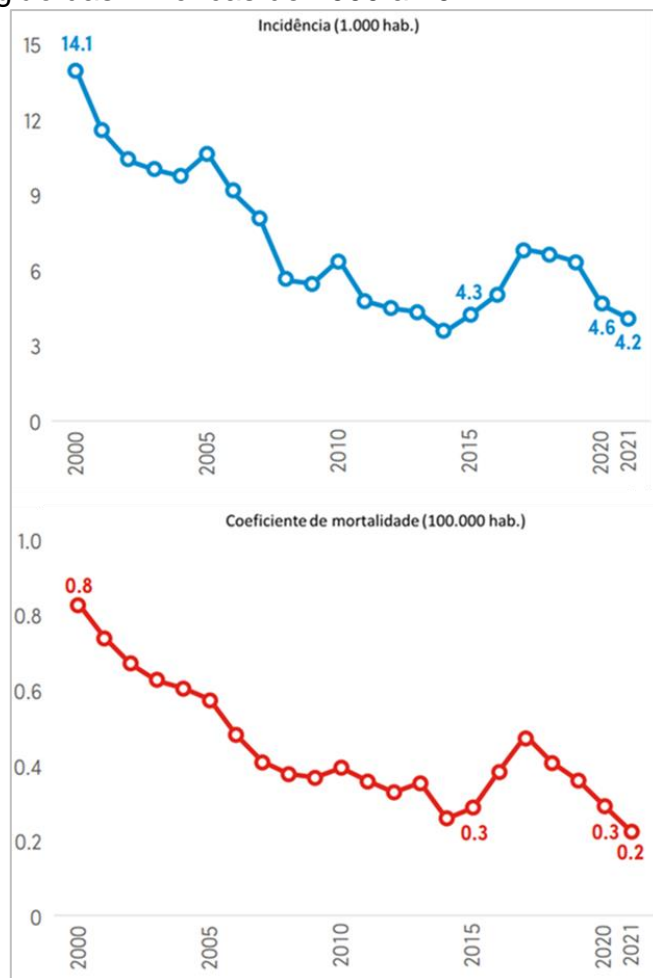
Fonte: Adaptado de World Malaria Report - OMS, 2022.

Em relação às estratégias e metas de eliminação da malária no mundo, havia uma expectativa de alcançar uma redução de 40% na incidência e na mortalidade até o ano de 2020, em comparação com os números de 2000 (OMS, 2015A). Além disso, esperava-se que a incidência fosse reduzida para

35 casos a cada 1.000 habitantes. No entanto, globalmente, as metas de eliminação não estão sendo alcançadas. É importante ressaltar que os objetivos de eliminação da malária em escala mundial têm como alvo uma redução de pelo menos 75% nos casos até 2025, em comparação com os números de 2015 (OMS, 2022).

Em 2021, na região das Américas foram registrados cerca de 600 mil casos e a incidência de 4 casos para cada mil habitantes. Cerca de 79% do total de casos está concentrado nos países na Venezuela, Brasil e Colômbia. Em 2021 foram registrados 334 óbitos por malária nas américas com coeficiente de mortalidade de 0,2 óbitos para cada 100 mil habitantes (OMS, 2022) (figura 3).

Figura 3. Incidência (1.000 hab.) e coeficiente de mortalidade de malária (100 mil hab.) na região das Américas de 2000 a 2021.



Fonte: Adaptado de World Malaria Report - OMS, 2022.

## *No Brasil*

Há relatos de que a urbanização desenfreada, o avanço acelerado da agricultura devastando as áreas naturais e a construção de cidades sem planejamento ecológico prévio contribuíram para a disseminação da malária em todo o mundo (RICH, 2009; HARPER, 2010).

No século XIX, foi identificado no Brasil a transmissão da malária por meio de mosquitos vetores (COX, 2002; TAUIL, 1985). No século XX, mais especificamente na década de 1930, relata-se o início dos registros de malária no Brasil, transmitida pelo vetor *Anopheles gambiae* (LOPES, 2019). Foi nessa década que a malária passou a ser considerada uma preocupação para a saúde pública (TAUIL, 1985). O número de casos de malária no país aumentou continuamente, levando à implementação de ações de controle da doença. Na década de 1940, foram notificados mais de 6 milhões de casos de malária anualmente (TAUIL, 2009). Até a década de 1960, observaram-se reduções nos níveis de malária, provavelmente como resultado das ações de controle vetorial, tratamento de pacientes sintomáticos e do desenvolvimento socioeconômico do Brasil, juntamente com o fortalecimento dos programas de controle da malária (OLIVEIRA-FERREIRA et al., 2010; TAUIL, 2009).

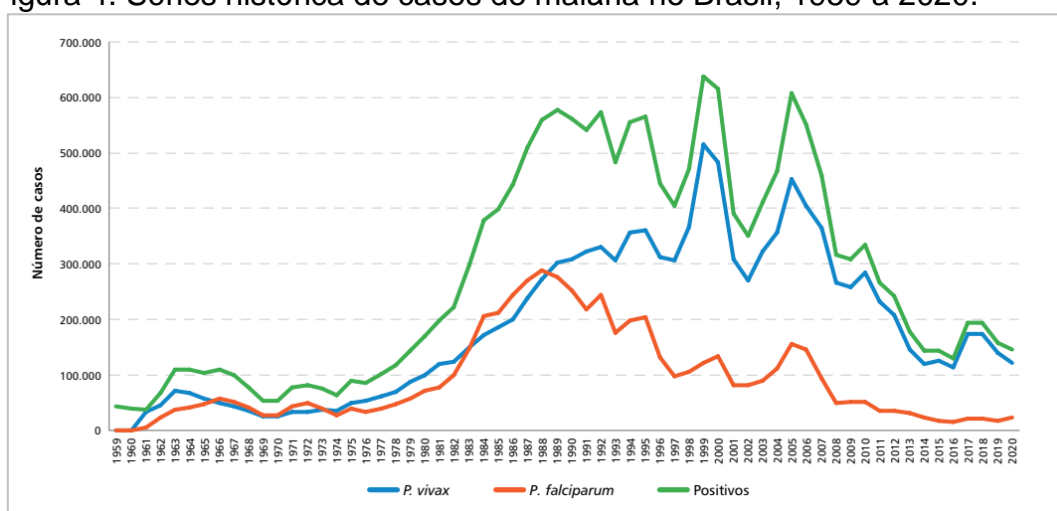
Entretanto, a partir de 1970, os casos de malária no Brasil registraram aumento ao tempo em que a transmissão passou a ocorrer por outras espécies de mosquitos vetores, como *Anopheles darlingi*, *Anopheles aquasalis* e *Anopheles albitarsis*. Em 1999, foram registrados cerca de 630 mil casos (MARQUES et al., 1994). Esse aumento acompanhou o crescimento rápido e desordenado de centros urbanos e a expansão da exploração da região Amazônica (AMZ) brasileira. O desenvolvimento urbano e demográfico resultou na criação de áreas propícias para a transmissão da doença, como áreas de mineração e garimpos (OLSON et al., 2010).

Com o aumento dos casos na região AMZ, o Ministério da Saúde (MS), seguindo as orientações da Organização Mundial da Saúde (OMS), passou a adotar estratégias de controle da malária que não se concentravam apenas no controle vetorial, mas também incluíam estratégias de controle integrado

envolvendo serviços de saúde e a população. Essas estratégias incluíram ações de educação em saúde, diagnóstico e tratamento oportuno (OMS, 1992; TAUIL, 2009).

Em 2021 foram registrado um total de 139.211 casos de malária (Figura 4), com uma redução de 4,1% em relação a 2020. O Brasil vem mostrando redução no total de casos de malária de forma constante desde 2005. Em 2017 e 2018 apresentou aumento no total de casos, mas voltou a controlar estes nos anos subsequentes, apesar dos casos de malária por *P. falciparum* e malária mista ter aumentado nos últimos 2 anos. Em 2021, o Brasil registrou um total de 1.354 casos importados de outros países, sendo quase toda sua totalidade proveniente de países vizinhos da América (BRASIL, 2022A).

Figura 4. Séries histórica de casos de malária no Brasil, 1959 a 2020.



Fonte: Boletim Epidemiológica Especial de Malária 2021 - Ministério da Saúde, 2021.

Em 2021 foram registrados 61 óbitos por malária (5 na região extra-AMZ), um aumento de 19,6% em relação a 2020, quando foram registrados 51 óbitos. Notou-se um aumento de aproximadamente 36,5% de 2020 para 2021 no total de óbitos nos estados da região AMZ brasileira. Na região extra-AMZ foi identifica uma redução de 50,0% no total de óbitos (BRASIL, 2023).

Considerando a Estratégia Técnica Mundial para a malária (ETG) 2016-2030 e o objetivo de eliminar a malária no Brasil, o país estabeleceu agendas

nacionais com o propósito de controlar, reduzir e, por fim, eliminar a doença até 2035. Essas agendas incluem as metas internas do Plano Plurianual (PPA), do Plano Nacional de Saúde (2020-2023) e do Programa de Qualificação das Ações de Vigilância em Saúde (PQA-VS). Cada uma dessas agendas tem a responsabilidade de monitorar aspectos específicos (BRASIL, 2021A).

A meta estabelecida para o PNS é de que até 2023 o Brasil tenha reduzido o número de casos autóctones para menos de 94 mil (BRASIL, 2021A). O plano nacional de eliminação da malária no Brasil visa uma redução para menos de 68 mil casos até 2025 e de 14 mil casos até 2030 e eliminação em 2035. A meta monitorada pelo PNS utiliza como referência os indicadores monitorados para o atingimento dos Objetivos de Desenvolvimento Sustentável (ODS) (BRASIL, 2022A).

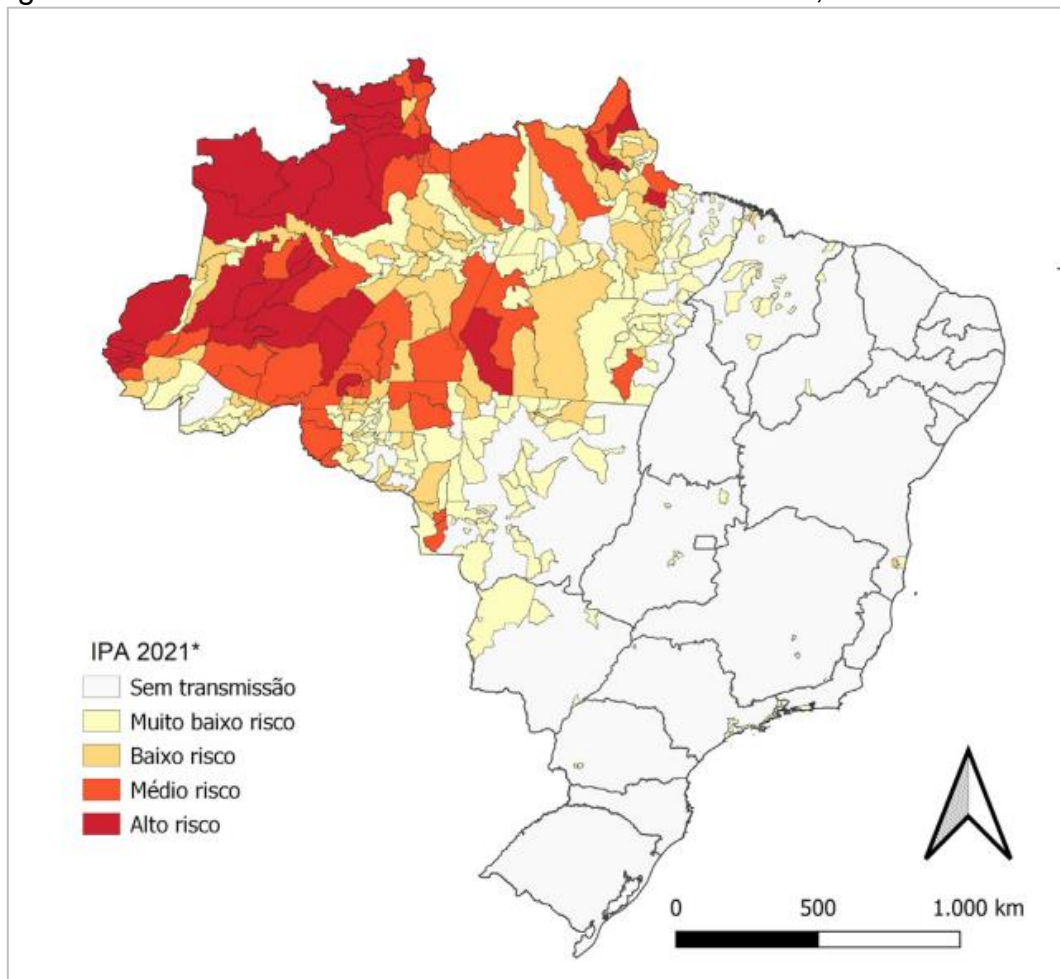
O objetivo 3.3 dos ODS visa acabar com epidemias de malária entre outras doenças infecciosas. Assim, o Brasil monitora a IPA para medir o risco de infecção pela doença no país.

Atualmente o Programa Nacional de Prevenção e Controle da Malária (PNCM) utiliza quatro classificações de risco com base na IPA:

- Muito baixo risco: IPA menor que 1 para cada 1.000 habitantes;
- Baixo risco: entre 1 a 10;
- Médio risco: entre 10 a 50, e;
- Alto risco: maior que 50.

Em 2021 foram registrados, 163 municípios (20,2% dos 808 municípios da região AMZ) com muito baixo risco, 61 (7,5%) com baixo risco, 38 (4,7%) com médio risco e 29 (3,6%) com alto risco de malária. Dos 808, 517 (64,0%) municípios não apresentaram casos autóctones (BRASIL, 2021A) (Figura 5).

Figura 5. Incidência Parasitária Anual de malária no Brasil, 2021.



Fonte: Boletim epidemiológico da malária – Ministério da Saúde, 2022.

## 1.2 Diagnóstico

O diagnóstico de casos de malária deve ser realizado oportunamente em até 24 horas, conforme as diretrizes estabelecidas pelo Ministério da Saúde. Esse diagnóstico oportuno tem um impacto significativo no início do tratamento, garantindo que o tratamento seja iniciado também de maneira oportuna e adequada. Isso, por sua vez, reduz o período de exposição do indivíduo infectado como fonte de transmissão local da doença. Além disso, é importante destacar que o diagnóstico oportuno e adequado desempenha um papel fundamental na prevenção de casos graves e na redução de possíveis

óbitos pela doença (WADI et al., 2016; BRASIL, 2021A).

O diagnóstico da malária pode ser realizado por meio de diversas técnicas, incluindo a gota espessa – que é entendida como padrão ouro para o diagnóstico por malária–, esfregaço delgado, testes de diagnóstico rápido (TDR) e técnicas moleculares (PCR). No Brasil, o Ministério da Saúde prioriza a técnica da gota espessa como principal método diagnóstico de malária, devido à sua alta acurácia, custo acessível e simplicidade de execução. A análise de amostras sanguíneas por microscopistas certificados e qualificados desempenha um papel crucial no sucesso das estratégias de controle, especialmente em regiões endêmicas para a malária. Essa técnica permite a identificação da espécie de plasmódio infectante e a determinação da parasitemia por volume sanguíneo ( $\text{mm}^3$ ).

Em áreas onde o acesso a microscopistas é difícil ou o transporte de lâminas é inviável, o Ministério da Saúde recomenda amplamente a utilização de TDR, o que é comum em regiões como os garimpos na região da Amazônia (MURTA et al., 2021). Técnicas moleculares para o diagnóstico da malária, embora eficazes na detecção de casos com baixa parasitemia, não são amplamente utilizadas devido à complexidade e ao alto custo envolvidos (BRASIL, 2021A).

No que diz respeito a regiões fora da Amazônia, é fundamental destacar que, devido à frequência de casos importados, as equipes de saúde devem permanecer alertas para a possibilidade de malária e conduzir investigações em casos suspeitos, a fim de garantir um diagnóstico oportuno e evitar o desenvolvimento de formas graves da doença (PINA-COSTA et al., 2009).

### **1.3 Tratamento**

Assim como o diagnóstico, o tratamento oportuno e adequado é uma atividade essencial a ser realizada pelos programas de vigilância e controle da malária, uma vez que a ausência de tratamento pode resultar em

complicações graves, incluindo complicações neurológicas, e potencialmente levar o paciente a óbito (BRASIL, 2020).

No Brasil, a disponibilização do diagnóstico e tratamento é gratuita e garantida pelo Sistema Único de Saúde (SUS). Diversos medicamentos e tratamentos estão disponíveis e suas indicações variam de acordo com a espécie de plasmódio infectante, a gravidade da doença, o histórico do paciente e também consideram características biológicas do paciente, como idade, peso e presença de condições associadas, como gravidez ou comorbidades (BRASIL, 2020; BRASIL, 2021A; FLANNERY et al., 2013). Há também recomendações específicas para pacientes que possuem deficiência de G6PD, uma enzima responsável pela proteção das hemácias contra lesões oxidativas que podem resultar em hemólise devido a tratamento por primaquina em infecções por *P. vivax* (MONTEIRO et al., 2014).

Os medicamentos indicados para o tratamento da malária no Brasil incluem a cloroquina, quinina, primaquina e derivados de artemisinina. A tafenoquina ainda está em fase de implementação no Brasil (SIQUEIRA et al., 2020; NEKKAB et al., 2020).

## **1.4 A Vigilância epidemiológica da malária**

### **1.4.1 Objetivos e características da Vigilância epidemiológica**

No Brasil, a vigilância epidemiológica da malária ocorre em ação conjunta das esferas municipais, estaduais e Federal de Governo tendo como referência o que é preconizado pelo PNCM. O Programa foi criado em 2003 sob competência da Secretaria de Vigilância em Saúde do Ministério da Saúde tendo como principais objetivos reduzir a mortalidade e incidência da doença, mantê-la ausente de áreas já sem transmissão e eliminá-la do Brasil (Brasil, 2022B).

Ainda, os objetivos da vigilância epidemiológica da malária no Brasil

são: Estimar a magnitude da morbidade e da mortalidade da doença, com identificação de grupos e áreas de risco para detecção precoce e controle de epidemias com interrupção de cadeias de transmissão (BRASIL, 2021A).

A vigilância epidemiológica da malária recomenda que, após um diagnóstico positivo, sejam imediatamente iniciadas ações de tratamento, investigação do local provável de infecção (LPI), busca ativa de casos e estratégias de controle vetorial, além de atividades de educação em saúde (BRASIL, 2022B).

Paralelamente às ações da vigilância epidemiológica, o controle e a prevenção da malária no Brasil também são influenciados pelas ações desenvolvidas pela vigilância entomológica. A vigilância entomológica fornece informações essenciais sobre o comportamento e a magnitude da população de vetores, auxiliando na gestão local. Além disso, essa área implementa intervenções específicas para o controle vetorial, como a utilização de substâncias químicas e a distribuição de mosquiteiros impregnados por inseticidas (BRASIL, 2022B).

As informações entomológicas e epidemiológicas se complementam, sendo ambas utilizadas para orientar as ações do PNCM. Além disso, a vigilância entomológica trabalha de forma integrada com a vigilância ambiental, considerando variáveis como as condições climáticas, características topográficas, áreas cobertas por matas ou que sofreram desmatamento, movimentos populacionais e o desenvolvimento de centros urbanos, bem como alterações nos corpos hídricos e áreas desmatadas, entre outros fatores (BRASIL, 2022B).

#### **1.4.2 Sistemas de Informação em Saúde para malária no Brasil**

As informações epidemiológicas de casos de malária são notificadas em Sistemas de Informação em Saúde (SIS) que são constantemente monitorados pelo Programa Nacional de Prevenção e Controle da Malária

(PNCM) e pelas demais esferas de gestão. Os casos de malária (CID-10: B50, B51, B52, B53, B54, P37.3 e P37.4) são de notificação compulsória e obrigatória de acordo com a portaria GM/MS Nº 420, DE 2 DE MARÇO DE 2022 do Ministério da Saúde (BRASIL, 2023B) e sua notificação é realizadas por meio de Fichas de notificação Individual (FNI) (ANEXO nº VII e VIII).

Para o monitoramento epidemiológico de malária são utilizados dois sistemas principais: O Sistema de Informação de Vigilância Epidemiológica da Malária (Sivep-Malária) e o Sistema de Informação de Agravos de Notificação (Sinan). O Sivep-Malária é responsável por receber notificações da região AMZ brasileira e o Sinan pelas notificações de casos de malária na região extra-AMZ.

Paralelamente ao Sivep-Malária há o Sistema de Informações e Controle de Vetores (Vetores-Malária) que contempla informações específicas para a vigilância entomológica e atividades de controle vetorial (NETO e CHIORO, 2021; REIS-SANTOS, 2023).

Anterior aos sistemas do Sivep-Malária e Sinan as notificações eram por formulários manuais até 1995 e entre 1996 e 2002 utilizou-se o Sistema de Informações do Programa Nacional de Controle Integrado da Malária (SISMAL) (BRASIL, 2002).

A notificação ocorre para casos suspeitos que venham a ser diagnosticados como positivos ou negativos. Ainda, registra-se quando o diagnóstico é para verificação de cura (Lâminas de Verificação de Cura - LVC). Os óbitos por malária são registrados no Sistema de Informação sobre Mortalidade (SIM).

O funcionamento destes sistemas, apesar de todos estarem sob gestão do Ministério da Saúde, não é integrado. Em 2021, iniciou-se o projeto de construção de um novo SIS que tem como objetivo integrar todas as informações presentes nos diferentes SIS sob gestão do MS, ou seja, integrar informação provenientes dos bancos de dados referentes a nascimentos, doenças de notificação compulsória e óbitos. O intuito deste projeto é unificar as informações dos usuários atendidos e cobertos pelo SUS o que facilitará o processo de vigilância em saúde (BRASIL, 2021A).

Enquanto este projeto está em fase de desenvolvimento, uma forma de conseguir integralizar informações de diferentes SIS para a investigação epidemiológica da malária é por meio de técnicas de vinculação de dados (GARCIA, MIRANDA e SOUSA, 2022).

Discussões sobre a necessidade de integralização destas informações já ocorriam no início do século (CAMARGO-JR e COELI, 2007; PINHEIRO et al., 2006). Além disso, mundialmente já se trabalhavam estudos e técnicas que pudessem concretizar a integralização de informações de diferentes SIS (ALMEIDA et al., 2020; OLVEIRA et al., 2016). Estas técnicas prezam, principalmente, pela vinculação de dados por meio de variáveis chaves de vinculação. Estas chaves são variáveis comuns entre os bancos, que na realidade brasileira podem ser exemplificadas com variáveis que contenham informações como Nome do paciente, Nome da mãe, Data de nascimento, Número do Cartão SUS, entre outras. Destaca-se que a utilização destas informações devem respeitar o que está disposto na lei que Regula o Acesso a Informações (Lei Nº 12.527, de 18 de novembro de 2011) e na Lei Geral de Proteção de Dados Pessoais (LGPD) (BRASIL, 2018).

### **1.4.3 Estratégias de controle da malária**

As principais estratégias de controle e prevenção da malária preconizadas pelo PNCM são as seguintes (BRASIL, 2022B):

Diagnóstico e tratamento oportuno: Realizados entre 24 e 48h após o início de sintomas, promovendo a cura do paciente, diminuindo o risco de complicações e interrompendo cadeias de transmissão (OMS, 2015B).

Controle vetorial: Estratégias de controle de vetores que ocorrem em parceria entre as áreas de vigilância epidemiológica, entomológica e ambiental, reduzindo o risco de transmissão da doença, diminuindo a morbimortalidade e interrompendo surtos da doença. Exemplos de estratégias utilizadas pelo PNCM incluem Controle Químico de Vetores Adultos,

Borrifação Residual Intradomiciliar, Mosquiteiros Impregnados com Inseticida de Longa Duração (MILDS), Nebulização Espacial e Controle Larvário (BRASIL, 2022B, BAIA-DA-SILVA et al., 2019).

Ações de educação em saúde: Atividades integradas entre os diferentes setores do SUS, como a vigilância epidemiológica, entomológica, ambiental e atenção à saúde. Essas ações impactam na educação continuada de profissionais de saúde e também na educação da população, expandindo seus conhecimentos sobre formas de prevenção da doença e como agir diante de uma possível infecção (BRASIL, 2022B).

Projeto apoiadores municipais da malária: Iniciado em 2012, é uma parceria entre o PNCM, a Fundação Oswaldo Cruz (Fiocruz) e a Fundação para o Desenvolvimento Científico e Tecnológico em Saúde (FIOTEC). O projeto é composto por profissionais multidisciplinares que atuam apoiando as gestões locais dos municípios prioritários para malária com o objetivo de fortalecer os serviços locais para o controle da malária (BRASIL, 2022B; GARCIA et al., 2023).

## **1.5 O plano de eliminação da malária no Brasil**

Os Objetivos de Desenvolvimento Sustentável (ODS) são uma série de metas globais estabelecidas pelas Nações Unidas para abordar uma variedade de questões relacionadas ao desenvolvimento sustentável até 2030. A malária não é especificamente mencionada como um ODS individual, mas está relacionada a vários dos ODS, especialmente ao ODS 3 - Saúde e Bem-Estar (RAVIGLIONE e MAHER, 2017).

Dentro do que tange o objetivo 3 dos ODS há o indicador 3.3 que busca "acabar com as epidemias de AIDS, tuberculose, malária e doenças tropicais negligenciadas até 2030" (RAVIGLIONE e MAHER, 2017). Deste modo a malária é abordada como parte dos esforços para combater doenças transmissíveis. Assim, o Brasil busca eliminar casos de malária autóctone no

país até 2035 (BRASIL, 2022C).

Aqui, então, é importante diferenciar os conceitos de eliminação, erradicação e controle: Eliminação de transmissão é quando se alcança zero casos autóctones transmitidos no local estudado; Erradicação é a redução permanente a zero casos como resultado de esforços deliberados, sem risco de reintrodução e; Controle é a redução da incidência, morbidade ou mortalidade da doença para um nível localmente aceitável fazendo necessárias medidas de intervenção contínuas para manter a redução (OMS, 2021B).

Eliminar completamente a malária no Brasil é um desafio complexo, mas o cumprimento do ODS 3.3, não necessariamente exige a eliminação total da malária em um país. O objetivo é reduzir significativamente a carga dessas doenças, tornando-as não mais uma epidemia (RAVIGLIONE e MAHER, 2017, BRASIL, 2022C).

No caso da malária, é possível trabalhar para reduzir o número de casos e mortes, tornando-a uma doença controlada em vez de uma epidemia descontrolada. Isso envolve a implementação de estratégias eficazes de prevenção, diagnóstico e tratamento da malária, bem como o fortalecimento dos sistemas de vigilância em saúde e a melhoria do acesso aos serviços de saúde para as populações em áreas de risco (BRASIL, 2022C).

Mesmo que a eliminação completa seja um objetivo ambicioso, o foco principal é garantir que a malária não seja mais uma ameaça significativa à saúde pública no Brasil até 2030. Alinhado a isso, reduzir os óbitos pela doença representará um avanço significativo no controle da doença, refletindo uma melhora no acesso e no diagnóstico oportuno (LAPORTA et al., 2022).

Com vistas a eliminar a malária no Brasil, o PNCM publicou em 2022 o Plano de eliminação da malária, em uma publicação nomeada como “Elimina malária Brasil”. Este, em consonância com o preconizado na ETG da OMS que estabelece como pilares a garantia de acesso universal à prevenção, diagnóstico e tratamento; a aceleração de esforços rumo à eliminação da malária e; a atribuição da vigilância da malária como uma intervenção essencial (BRASIL, 2022C).

O Plano Nacional de Eliminação busca reduzir os casos de malária para menos de 68 mil até 2025 e menos de 14 mil até 2030, com o objetivo final de eliminação até 2035. O plano é dividido em quatro fases: preparação, consolidação da redução, eliminação e prevenção de retransmissão. O planejamento local levará em conta o diagnóstico situacional e o risco de malária em cada município. Apesar dos desafios da COVID-19, o Brasil reafirma seu compromisso em reduzir e eliminar a malária, envolvendo diversos setores da sociedade (BRASIL, 2022C).

## **1.6 Vinculação de bancos de dados**

A falta de interoperabilidade entre sistemas de saúde, devido à ausência de uma chave única de identificação de pacientes, como CPF ou RG, requer o uso de técnicas probabilísticas de vinculação de dados, que avaliam a similaridade entre variáveis identificadoras, como nome e data de nascimento (GARCIA, MIRANDA e SOUSA, 2022). Apesar de sua dependência de recursos especializados e dados de alta qualidade, essas abordagens têm demonstrado sucesso em pesquisas em Vigilância em Saúde (OLIVEIRA et al., 2016, LIMA et al., 2020).

Os métodos de vinculação probabilística de dados têm se mostrado uma solução viável para integrar informações de diferentes sistemas de saúde (OLIVEIRA et al., 2016), utilizando um conjunto de regras para determinar a similaridade entre registros e identificar pares verdadeiros (GARCIA, MIRANDA e SOUSA, 2022).

A vinculação de dados identifica pares de registros que concordam a partir de um determinado conjunto de regras (BRUSTULIN e MARSON, 2018). É indicado quando os bancos de dados que serão trabalhados apresentem uma variável identificadora comum ou um conjunto de variáveis com boa qualidade de preenchimento (MACHADO, 2004; QUEIROZ et al., 2010; PERES, et al., 2016).

No processo de vinculação de dados é essencial a utilização de variáveis chaves que possuam aspectos de identificação do usuário como “nome do paciente”, “data de nascimento” e “nome da mãe”, que são chaves recorrentemente utilizadas na literatura (CHURCHES et al., 2002; YAOHAO e MATION, 2018).

Para o melhor aproveitamento do processo de vinculação é necessário que sejam feitos padronizações e tratamento das variáveis (GARCIA, MIRANDA e SOUSA, 2022).

Este processo de padronização das variáveis de nome do notificado, CPF, data de nascimento e nome da mãe considera a possibilidade de pequenas inconsistências no preenchimento destas informações pelos notificantes nos diferentes sistemas de informação (GARCIA, MIRANDA e SOUSA, 2022).

Após a limpeza e padronização destas informações, pode-se explorar o banco de dados para verificação de registros duplicados. Porém, a exclusão ou não destes registros duplicados devem ser discutidos antecipadamente considerando o objetivo do resultado da vinculação de dados (GARCIA, MIRANDA e SOUSA, 2022).

Previamente à vinculação também devem ser estabelecidos quantas chaves serão utilizadas no processo. Pois a combinação de diferentes chaves, como a chave “Nome do paciente” combinada com a chave “Data de nascimento” acarretará um processo de vinculação mais específico, prezando por maior número de registros verdadeiros-positivos. A utilização de apenas uma chave, como “Nome do paciente” apresentará resultados mais sensíveis, com maior presença de pareamentos falsos-positivos (GARCIA, MIRANDA e SOUSA, 2022).

A decisão de se utilizar mais de uma chave para a vinculação deve ser embasada com análises de qualidade dos bancos de dados, pois o processo determinístico pode ser afetado por dados digitados incorretamente, por exemplo, datas de nascimentos que são divergentes, mas que em análises manuais o pesquisador consegue identificar que são as mesmas datas, mas que foram inseridas com algum erro de digitação. Dessa forma, pode ser

interessante utilizar métodos mais sensíveis desde que acompanhados por limpezas manuais de possíveis pareamentos verdadeiros-negativos. Porém, processos de revisão manual nem sempre são viáveis (GRANNIS et al., 2003).

## **1.7 Modelos de predição temporal**

Uma estratégia de monitoramento do alcance das metas de eliminação é por meio de utilização de modelos preditivos de séries temporais (*forecast analysis*) que já são amplamente utilizados em estudos epidemiológicos (LIMA; LAPORT, 2021; LAPORTA et al., 2022; FIKRY; INOUW, 2023) e se mostram como bons atributos para otimizar o monitoramento e vigilância epidemiológica da malária.

Um dos métodos de predição temporal popular é o modelo de suavização e predição exponencial de Holt-Winters (HW) que permite a contabilização em seus modelos das influências das tendências, sazonalidades e do efeito aleatório da série temporal (KOEHLER, SNYDER e ORD, 2001; SWAPNAREKHA et al., 2021). A suavização exponencial usa uma média móvel ponderada exponencialmente (EWMA) para nivelar a série temporal e fornecer uma previsão de maior qualidade. O modelo HW multiplicativo sazonal de HW multiplica a tendência da série pela sazonalidade, em vez de somar uma à outra, como ocorre na modelagem aditiva de HW. O modelo multiplicativo é mais adequado para séries temporais de malária porque sua sazonalidade tem maior influência no aumento ou diminuição do número de casos. A previsão dos cenários futuros é acompanhada por estimativas de intervalos de confiança de 95%. Ainda, testes de adequação do modelo podem ser realizados KOEHLER, SNYDER e ORD, 2001; SWAPNAREKHA et al., 2021).

## 2. JUSTIFICATIVA

A justificativa para a realização deste estudo é pautada na preocupação decorrente da persistência de casos e óbitos por malária no Brasil, apesar da robusta infraestrutura de vigilância epidemiológica e da disponibilidade de testes diagnósticos e tratamentos gratuitos. Ainda que o país tenha avançado significativamente no controle da malária, a presença contínua de óbitos sugere a existência de fatores complexos e variáveis associadas, como condições de saúde pré-existentes e tratamentos não oportunos, que precisam ser investigados e compreendidos em profundidade.

A compreensão do cenário epidemiológico da malária no Brasil torna-se imperativa, uma vez que esse conhecimento pode lançar luz sobre as circunstâncias e determinantes que influenciam a ocorrência desses óbitos. Identificar e analisar esses fatores é fundamental para aprimorar as estratégias de prevenção e tratamento da malária, bem como direcionar recursos e esforços de forma mais eficaz. Além disso, essa pesquisa pode contribuir para o direcionamento de intervenções específicas que visam reduzir a mortalidade relacionada à malária.

É importante ressaltar que, embora esta pesquisa não ofereça benefícios diretos a indivíduos, ela visa um impacto substancial em nível coletivo. Os resultados fornecerão evidências sólidas às organizações públicas de saúde do Brasil, auxiliando na formulação de políticas públicas baseadas em evidências. Essas políticas, por sua vez, têm o potencial de gerar melhorias significativas na saúde da população brasileira, especialmente daqueles que são mais vulneráveis à malária. Portanto, o estudo desempenha um papel crucial na promoção da saúde pública e na mitigação do impacto da malária no país.

### **3. OBJETIVOS**

#### **3.1. Objetivo geral**

Analisar o cenário epidemiológico da malária no Brasil e investigar os fatores associados às mortes por essa doença no país entre 2011 e 2020.

#### **3.2. Objetivos específicos**

1. Descrever o cenário epidemiológico de casos de malária na região extra-Amazônica brasileira de 2011 a 2020.
2. Descrever o cenário epidemiológico de casos de malária na região Amazônica brasileira de 2011 a 2022.
3. Descrever processos metodológicos de vinculação de bancos de dados (*record linkage*) para sistemas de informação em saúde de malária.
4. Analisar os fatores associados às mortes por malária no Brasil de 2011 a 2020.

## **4. MATERIAIS E MÉTODOS**

### **4.1. Aspectos metodológicos gerais**

Esta tese é composta por uma coletânea de pesquisas científicas com diferentes metodologias de epidemiologia descritiva e analítica. Neste trabalho constam pesquisas com a utilização de metodologias de análises de séries temporais, vinculação de dados e um estudo caso-controle. Analisou-se os cenários epidemiológicos a partir de 2011 para se trabalhar com os contextos epidemiológicos mais recentes do Brasil e assim tecer discussões mais aplicáveis à situação atual do país.

### **4.2. Aspectos metodológicos específicos**

#### **4.2.1. Descrição cenário epidemiológico da malária na região extra-Amazônica brasileira (Objetivo específico 1)**

##### *Delineamento do estudo*

Estudo Ecológico de Séries Temporais de casos e óbitos por malária notificados na região extra-AMZ do Brasil entre 2011 e 2020.

##### *Área, População e Período de estudo*

Foram adotadas as seguintes definições para casos de malária: Caso notificado – Caso com resultado laboratorial positivo para malária; nova infecção (caso novo) – Caso com resultado laboratorial positivo para malária excluindo lâmina de verificação de cura (LVC) positiva; Casos autóctones – Classificação com base no local de provável infecção (município ou estado); Casos negativos não foram analisados neste estudo. As mortes por malária foram analisadas de acordo com o local de residência do caso.

O local de estudo selecionado foi a região extra-AMZ brasileira que abrange 17 estados (Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe, Bahia, Minas Gerais, Espírito Santo, Rio de Janeiro, São Paulo, Paraná, Santa Catarina, Rio Grande do Sul, Mato Grosso do Sul, Goiás) e Distrito Federal. Ressalta-se que a Amazônia Legal abrange todos os estados da região Norte mais os estados do Maranhão, Mato Grosso e Tocantins. Na região extra-AMZ, segundo projeções populacionais do Instituto Brasileiro de Geografia e Estatística (IBGE), há uma média de 176.163.089 habitantes por ano, entre 2011 e 2020.

#### *Fonte e análise de dados*

Os dados públicos do Sistema de Informação de Agravos de Notificação (Sinan) foram obtidos no portal do banco de dados do Ministério da Saúde – DATASUS – em maio de 2021. Governo Federal em junho de 2021.

#### *Análise Estatística*

Para processamento e análise dos dados, utilizou-se o software Qgis Desktop (2.18), Microsoft Excel (2016) e os pacotes read.dbc, tidyverse, lubridate, forecast e fpp do software R (versão 4.0. 5).

Foram analisadas as variáveis relacionadas à data de notificação, data de início dos sintomas, data de início do tratamento, data de nascimento, idade, sexo, raça/cor, escolaridade, gestante, atividade nos últimos 15 dias (ocupação), atividade profissional, tipo de lâmina (ativa, detecção passiva, lâmina de verificação de cura - LVC), resultados de exames, autoctonia, locais de notificação – Unidades da Federação (UF) e Município e local do óbito.

Taxas de incidência de malária (Total de casos novos de malária por município de provável infecção/População total do município \* 100.000) e taxas de notificação de malária (Total de casos de malária por município de notificação/População total do município \* 100.000) foram calculados para cada ano. As taxas de incidência calculadas pelo PI consideram que esses casos são autóctones. O número total de novas infecções é resultado do total

de casos positivos no período excluindo aqueles classificados como LVC - que representam recorrências da infecção.

#### *Análise descritivas espaciais*

O plugin do software Qgis "Flowmaps" (versão 2.18) foi utilizado para analisar o fluxo entre PI e o local de notificação de novos casos de malária. Juntamente com a análise de fluxo, analisamos a taxa de incidência média para o período.

#### *Análises de Previsão*

Para prever casos para 2021 e 2022 a partir da análise de séries temporais foram utilizados os comandos `ts`, `decompose`, `Holt-Winters` e `forecast` no software R. Para a análise de previsão, foram considerados os atributos de média, tendência e sazonalidade dos casos notificados de malária de 2011 a 2020. Usamos métodos de suavização e predição exponencial de Holt-Winters (HW) com influência multiplicativa da sazonalidade na modelagem, como usado no estudo de Swapnarekha et al (2021).

#### *Considerações Éticas*

Este estudo não necessitou de aprovação por comitês de ética seguindo a resolução do Conselho Nacional de Saúde nº 446/2011.

### **4.2.2. Descrição do cenário epidemiológico da malária na região Amazônica brasileira (Objetivo específico 2)**

#### *Desenho do estudo*

Este estudo utilizou um desenho descritivo com análise de séries temporais para examinar as tendências nas notificações de casos de malária na região AMZ brasileira de 2011 até outubro de 2022. Os dados utilizados foram registros não-nominais obtidos do Sistema de Informação Epidemiológica da Malária no Brasil (Sivep-Malária). Os dados populacionais foram obtidos do Instituto Brasileiro de Geografia e Estatística (IBGE).

### *Local do Estudo e População*

O estudo concentrou-se na região AMZ brasileira, que abrange nove estados: Acre, Amazonas, Amapá, Mato Grosso, Maranhão, Pará, Roraima, Rondônia e Tocantins. Estes estados englobam 808 municípios, cobrindo uma área de mais de 5 milhões de km<sup>2</sup>, correspondendo a aproximadamente 58,9% da área total do país. O banco de dados do Sivep-Malária inclui notificações de todos os nove estados, que tiveram uma população estimada de 28,4 milhões em 2022 (IBGE, 2023). A população do estudo compreendeu indivíduos notificados no Sivep-Malária como novos casos de infecção por malária entre janeiro de 2011 e outubro de 2022 (variáveis de filtro: ID\_LVC = 2 e TIPO\_LAM = 1 e 2). Os casos autóctones foram classificados com base em seu local provável de infecção, seguindo as diretrizes fornecidas pelo MS (BRASIL, 2022B; OMS, 2019).

### *Variáveis do Estudo*

A malária no Sivep-Malária é notificada como um novo caso ou como uma recorrência - lâmina de verificação de cura. O estudo analisou a data das notificações e tratamento, resultados laboratoriais que definem a espécie de *Plasmodium*, sexo, idade, escolaridade, raça/cor, gravidez, presença de sintomas, tipo de detecção, parasitemia, atividades nos últimos 15 dias, local provável de infecção (país, estado, municípios) e tipo de local provável de infecção.

### *Análise Estatística*

A análise descritiva envolveu a sumarização do número anual de casos de malária. Devido à baixa proporção de malária causada por *P. malariae* e *P. ovale*, esses casos foram analisados juntamente com infecções por *Plasmodium vivax* (*Pv*). Da mesma forma, infecções de espécies mistas que incluíam infecções por *Plasmodium falciparum* (*Pf*) foram categorizadas como infecções por *Pf*. As taxas de incidência (IR) foram calculadas dividindo o

número total de novos casos pela população estimada de cada município e multiplicando-a por 1.000 residentes, representando a população em risco (IPA - incidência parasitária anual). As taxas de incidência de malária foram categorizadas com base na estratificação de risco fornecida pelo Ministério da Saúde do Brasil (BRASIL, 2021A; BRASIL, 2022A) classificando a IR como alto risco (IR  $\geq$  50 casos por 1.000 residentes), médio risco (IR entre 10 e 49,99 casos por 1.000 residentes), baixo risco (IR entre 1 e 9,99 casos por 1.000 residentes) e muito baixo risco (IR  $<$  1 caso por 1.000 residentes).

### *Análises de Tendência*

Descrevemos as tendências dos casos clínicos de malária ao longo do tempo usando o software Joinpoint Regression Versão 4.9.1.0. A análise utilizou modelos de regressão de Poisson para identificar pontos de inflexão (joinpoints) e Mudanças Percentuais Mensais (MPC) para determinar as tendências (aumento, diminuição ou estabilidade (KIM et al., 2000). Um nível de significância inferior a 0,05 foi considerado para todas as análises estatísticas de tendência. O modelo de Regressão Joinpoint testa se uma linha de múltiplos segmentos fornece um ajuste estatisticamente melhor para descrever a evolução temporal dos dados em comparação com uma linha reta ou uma linha com menos segmentos (KIM et al., 2000). O método de Busca em Grade é empregado para identificar a melhor adequação dos dados para a série temporal, e permutações de Monte Carlo são usadas para testar e comparar diferentes números de pontos de junção em cada modelo (KIM et al., 2000; BANDI et al., 2015).

### *Análises de Previsão*

Os pacotes forecast e fpp foram utilizados para prever os casos de malária para 2023 e 2024. Utilizou-se o método Holt-Winters que é adequado para previsões de curto prazo. Para verificar a adequação do modelo, a mesma análise foi realizada para prever casos para 2022, e as observações reais e os valores previstos mostraram resultados semelhantes. Os modelos de previsão abrangeram todos os casos autóctones de malária na região AMZ

brasileira, bem como análises específicas para infecções por *P. falciparum*, infecções mistas, estados individuais, aldeias de população indígena e áreas de garimpo.

#### *Análises Descritivas Espaciais*

O software Qgis foi utilizado para gerar a distribuição de casos de acordo com o município de infecção com base em suas classificações de risco para cada ano. Foram realizadas duas análises espaciais: uma para todos os casos autóctones de malária e outra para casos autóctones de *P. falciparum* e infecções mistas. A análise foi conduzida para descrever o comportamento espacial das taxas de incidência de malária na região AMZ.

#### *Considerações Éticas*

Este estudo foi isento de aprovação por comitês de ética, seguindo as diretrizes estabelecidas pela Resolução nº 466/2012 do Conselho Nacional de Saúde do Brasil. Os dados foram fornecidos pelo Ministério da Saúde do Brasil (processos #25072.015038/2021-05, #25072.034317/2021-60 e #25072.004310/2023-85).

### **4.2.3. Descrição dos aspectos metodológicos de *record linkage* para sistemas de informação de malária (objetivo 3)**

#### *Delineamento de estudo*

Estudo metodológico com caráter instrucional sobre os procedimentos de limpeza e tratamento de dados que precedem a vinculação (*Record linkage*) e os comandos utilizados para executar a técnica de linkage utilizando o software R e RStudio.

#### *Preparação dos bancos de dados*

É importante que, antes de iniciar o processo de vinculação de dados, sejam definidas as variáveis que serão utilizadas para a análise dos dados

vinculados. Esta definição subsidiará o processo de tratamento dos dados, pois ao se trabalhar com bancos de dados de grande magnitude pode ser necessário diminuir o tamanho do banco para que a máquina (computador) possa processar os dados com maior facilidade evitando problemas de performance como travamento ou encerramento do programa de forma indesejada.

#### *Vinculação dos bancos de dados*

Para a vinculação dos bancos de dados do SIM com o Sivep-Malária e com o Sinan foram utilizadas as variáveis chaves “Nome do paciente”, “Nome da mãe” e “Data de nascimento”. As variáveis passaram por tratamento prévio, com remoção de caracteres inconsistentes como pontuações gráficas, números, acentos gráficos e palavras conectoras.

Utilizou-se o software R para realização da vinculação. Neste, utilizou-se o pacote “*fuzzy join*” que permite a vinculação de dois bancos de dados diferentes por meio de técnicas probabilísticas que verificam a dissimilaridade entre as o texto de cada variável. Aqui, utilizou-se os métodos de Cosine e Leveshtein (ZHANG; HU; BIAN, 2017; LIAO; XU, 2015).

#### *Exclusão de falsos-positivos*

Devido ao número elevado de registros pós-vinculação e a impossibilidade de revisão manual estabeleceu-se uma regra de validação que considerou como verdadeiros-positivos apenas os pares em que o conteúdo de pelo menos uma das três variáveis-chaves fosse idêntico.

#### *Considerações éticas*

Este projeto recebeu aprovação do comitê de ética em pesquisa da Faculdade de Medicina da Universidade de Brasília em janeiro de 2022. A aprovação está documentada sob o protocolo CEP: 5.216.688 (51246121.0.0000.5558).

#### **4.2.4. Descrição dos fatores associados ao óbito por malária (objetivo 4)**

##### *Delineamento do estudo*

Estudo do tipo caso-controle com utilização de dados secundários. A população de estudo será de casos de malária e óbitos, registrados no Sinan, Sivep-Malária e SIM.

##### *Área, População e Período de estudo*

Foram analisados óbitos ocorridos em todo o território brasileiro, incluindo região da Amazônia legal que é composta pelos estados do Acre, Amapá, Amazonas, Mato Grosso, Maranhão, Pará, Rondônia, Roraima e Tocantins e demais estados e Distrito Federal na região extra-AMZ. O período análise será de 2011 a 2020.

##### *Fonte dos dados*

Para realização deste estudo fora solicitado acesso a diferentes bancos de dados sob supervisão do Ministério da Saúde do Brasil descrito a seguir. Para construção do banco de dados de estudo fora solicitado acesso aos dados provenientes dos seguintes sistemas de informação: Sistema de vigilância epidemiológica da malária (Sivep-Malária) para casos positivos de malária registrados na região da Amazônia legal; Sistema de Informação de Agravos de Notificação (Sinan) para casos positivos de malária na região extra-AMZ; Sistema de Informação de Mortalidade (SIM) para registros de óbitos por malária (CID-10: B50 a B54).

##### *Definição de Casos e Controles*

As notificações de óbitos individuais que tiverem pareamento de dados com os bancos de dados do Sivep-Malária ou Sinan foram adicionadas no estudo como notificações de Casos. Os óbitos sem pareamento não foram incluídos no estudo. As notificações de casos de malária presentes no Sivep-Malária e Sinan que não tiverem pareamento com o banco de dados do SIM

foram incluídas no estudo como notificações de Controles. Notificações individuais de LVC não foram incluídas como controle.

Para exclusão dos possíveis resultados falsos-positivos, utilizou-se uma combinação de fatores onde pelo menos uma das três chaves utilizadas (nome do paciente, nome da mãe, data de nascimento) chaves de vinculação utilizadas tenham tido um resultado 100% igual entre as bases de dados. A conferência e exclusão manual dos falsos-positivos não foi possível devido a inacessibilidade ao dado nominal.

### *Análises estatísticas*

O banco de casos e controles fora explorado e caracterizado com base em análises descritivas com utilização de médias, medianas, mínimos, máximos, inter-quartis e desvios-padrões. Foram quantificados indicadores epidemiológicos como taxas de mortalidade e taxas de letalidade. Para o cálculo das taxas de mortalidade utilizou-se dados de estimativas populacionais extraídos do IBGE para o período de 2011 a 2020 como denominador.

Para a definição de fatores de risco e causas associadas foram calculadas as razões de chances (Odds-Ratio) entre as variáveis de interesse. As variáveis que sinalizaram associação por meio do teste de qui-quadrado foram inseridas em modelos de análises de regressão logística. Para a regressão logística fora utilizada o método de stepwise incluindo variáveis uma a uma para verificar o melhor comportamento do modelo estatístico (SARKAR et al., 2009; MEIRELES et al., 2020).

### *Considerações éticas*

Este projeto recebeu aprovação do comitê de ética em pesquisa da Faculdade de Medicina da Universidade de Brasília em janeiro de 2022. A aprovação está documentada sob o protocolo CEP: 5.216.688 (51246121.0.0000.5558).

## 5. RESULTADOS

### 5.1. Descrição do comportamento e cenário epidemiológico da malária na região extra-Amazônica brasileira (Objetivo específico 1)

#### 5.1.1. Série temporal da malária extra-AMZ no Brasil: cenário epidemiológico e um modelo de previsão de dois anos. (\*Artigo publicado no Malaria Journal - Apêndice II)

## INTRODUÇÃO

A malária é uma doença infecciosa febril aguda, causada por um protozoário do gênero *Plasmodium* que infecta humanos através da picada do mosquito vetor fêmea do gênero *Anopheles*. É frequente em países tropicais e em desenvolvimento. De acordo com a Organização Mundial da Saúde (OMS, a malária afeta cerca de 212 milhões de pessoas em todo o mundo e causou mais de 409.000 mortes em 2019, com taxas de mortalidade variando de 0,2 a 2,2 em todo o mundo [1,2].

No Brasil, as principais espécies de plasmódios são *P. vivax* e *P. falciparum*, há também a transmissão de *P. malariae* mas sua ocorrência é baixa em comparação com a de *P. vivax* e *P. falciparum*. A transmissão da malária em áreas endêmicas e não endêmicas dá-se por mosquitos vetores do gênero *Anopheles*, uma espécie comum em áreas de Floresta Amazônica e Mata Atlântica [3-6].

Historicamente, a malária no Brasil passou por mudanças em sua distribuição geográfica, onde sua ocorrência ao longo dos 20<sup>o</sup> século foi predominante em todo o Brasil e nos últimos 50 anos se concentrou na região AMZ brasileira [7,8]. A região da Amazônia Legal compreende aproximadamente 99% de todos os casos do país. Embora a região extra-

AMZ represente apenas 1% do total de casos, vale ressaltar que os casos nesta região apresentam maior letalidade e possível restabelecimento da transmissão da malária [9]. De 2011 a 2020, Brasil registrou cerca de 1,8 milhão de casos de malária com 450 óbitos [10].

Casos de malária na região extra-AMZ têm frequência de óbitos menor do que na Amazônia legal, porém, letalidade de casos é 123 vezes maior do que na região AMZ em 2019 [3,11].

Assim, este estudo tem como objetivo descrever e caracterizar o cenário e perfil epidemiológico da malária na região extra-AMZ do Brasil nos últimos 10 anos.

## **MÉTODOS**

### *Desenho de estudo*

Estudo ecológico de séries temporais com a descrição de casos e óbitos por malária notificados na região extra-AMZ do Brasil de 2011 a 2020.

Neste estudo, foram adotadas as seguintes definições para casos de malária: Caso notificado – Caso com resultado laboratorial positivo para malária; nova infecção (caso novo) – Caso com resultado laboratorial positivo para malária excluindo lâminas de verificação de cura (LVC) positiva; Casos autóctones – Classificação com base no local de provável infecção (município ou estado). Casos negativos não foram analisados neste estudo. As mortes por malária foram analisadas de acordo com o local de residência do caso [12,13].

### *Local de estudo*

O local de estudo selecionado foi a região extra-AMZ brasileira que abrange 17 estados (Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe, Bahia, Minas Gerais, Espírito Santo, Rio de Janeiro, São Paulo, Paraná, Santa Catarina, Rio Grande do Sul, Mato Grosso do Sul, Goiás) e Distrito Federal. Ressalta-se que a Amazônia Legal abrange

todos os estados da região Norte mais os estados do Maranhão, Mato Grosso e Tocantins [14]. Na região extra-AMZ, segundo projeções populacionais do Instituto Brasileiro de Geografia e Estatística (IBGE), há uma média de 176.163.089 habitantes por ano, entre 2011 e 2020.

#### *Fonte de dados e análise*

Os dados públicos do Sistema de Informação de Agravos de Notificação (Sinan) foram obtidos no portal do banco de dados do Ministério da Saúde (MS) – DATASUS – em maio de 2021. Para processamento e análise dos dados, utilizou-se o software Qgis Desktop (2.18), Microsoft Excel (2016) e os pacotes read.dbc, arrumaverse, lubridate, forecast e fpp do software R (versão 4.0. 5) [15].

Foram analisadas as variáveis relacionadas à data de notificação, data de início dos sintomas, data de início do tratamento, data de nascimento, idade, sexo, raça/cor, escolaridade, gestante, atividade nos últimos 15 dias, atividade profissional, tipo de lâmina (ativa, detecção passiva, Lâmina de verificação de cura - LVC), resultados de exames, autoctonia, locais de notificação – Unidades da Federação (UF) e Município, Local de provável infecção (PI) e local do óbito.

Taxas de incidência: Taxas de incidência de malária (Total de casos novos de malária por município de PI/População total do município\*100.000) e taxas de notificação de malária (Total de casos de malária por município de notificação/População total do município\*100.000) foram calculados para cada ano. As taxas de incidência calculadas pelo PI consideram que esses casos são autóctones. O número total de novas infecções é resultado do total de casos positivos no período excluindo aqueles classificados como lâminas de verificação de cura (LVC) - que representam recorrências da infecção [12].

O plugin do software Qgis "Flowmaps" (versão 2.18) foi utilizado para analisar o fluxo entre PI e o local de notificação de novos casos de malária. Juntamente com a análise de fluxo, analisamos a taxa de incidência média para o período (Média de: Total de novas infecções por município de PI por ano/ População residente por município por ano x 100.000 habitantes).

Para prever casos para 2021 e 2022 a partir da análise de séries temporais foram utilizados os comandos `ts`, `decompose`, `Holt-winters` e `forecast` no software R. Para a análise de previsão, foram considerados os atributos de média, tendência e sazonalidade dos casos notificados de malária de 2011 a 2020. Usamos métodos de suavização e predição exponencial de Holt-winters com influência multiplicativa da sazonalidade na modelagem, como usado no estudo de Swapnarekha et al [16]. O modelo multiplicativo sazonal de HW multiplica a tendência da série pela sazonalidade, em vez de somar uma à outra, como ocorre na modelagem aditiva de HW. O modelo multiplicativo é mais adequado para séries temporais de malária porque sua sazonalidade tem maior influência no aumento ou diminuição do número de casos. Se a sazonalidade não tivesse uma influência tão alta na série temporal, o modelo aditivo seria a melhor escolha para fazer a modelagem de previsão.

O modelo de previsão de Holt-Winters avalia o comportamento das séries temporais com base nos padrões dos três atributos mencionados anteriormente. A suavização exponencial usa uma média móvel ponderada exponencialmente (EWMA) para nivelar a série temporal e fornecer uma previsão de maior qualidade. Este modelo usa suavização exponencial para considerar valores históricos e prever um valor futuro provável considerando intervalos de confiança de 95% [17].

Este estudo não necessitou de aprovação por comitês de ética seguindo a resolução do Conselho Nacional de Saúde nº 446/2011.

## **RESULTADOS**

De 2011 a 2020, a região extra-AMZ brasileira registrou 6.849 notificações de casos confirmados de malária, foram 682 (9,9%) casos classificados como LVC e 6.167 (90,1%) foram notificações de casos novos de malária, onde 1.640 (26,5%) foram notificados por meio de detecção ativa.

Os casos predominaram no sexo masculino (78%), brancos (44,3%) e pardos (38,7%); a distribuição por raça/cor é semelhante entre homens e mulheres. A maior concentração por faixa etária está em pessoas entre 20 e 59 anos de idade (84%). A faixa etária de 30 a 39 anos apresenta a maior frequência de casos (26,1%) com brancos ou pardos (42,8% e 38%, respectivamente). Entre os casos positivos, foram encontrados 58 casos com idade auto referida menor que um ano, com mediana de 37 anos (1-quartil: 28; 3-quartil: 49; máximo: 93). Quanto à escolaridade, as maiores frequências estão em pessoas com 9 a 11 anos de escolaridade (21,7%). Para “nível de escolaridade”, 23,6% foram preenchidos “ignorados” (Tabela 1).

Quanto às ocupações, as atividades mais frequentes estão relacionadas a “viagens” e “turismo”, que respondem por 31,2% do total de registros de casos de malária na região extra-AMZ. As mulheres apresentam maior frequência relativa de atividades relacionadas a "turismo" e empregados "domésticos", enquanto os homens apresentam maior frequência relativa de casos nas atividades de "agricultura", "construção" e "mineração" (Tabela 1). Analisando as profissões de acordo com o nível de escolaridade, as mais frequentes são “Viajante” e “Turismo”, com frequência de 4,6% e 2,6% quando acima de 12 anos de estudo.

Tabela 1. Perfil epidemiológico dos casos de malária na região extra-AMZ do Brasil, 2011 a 2020.

	<b>Sexo</b>					
	<b>Masculino</b>		<b>Feminino</b>		<b>Total</b>	
	n	%	n	%	n	%
<b>Total</b>	5.343	78,01	1.506	21,99	6.849	100
<b>Faixa etária</b>						
< 1 ano	45	0,84	14	0,93	59	0,86
01/abr	52	0,97	29	1,93	81	1,18
05/set	37	0,69	24	1,59	61	0,89
out/14	85	1,59	54	3,59	139	2,03
15 - 19	180	3,37	69	4,58	249	3,64
20 - 29	1.096	20,51	357	23,71	1.453	21,21
30 - 39	1.461	27,34	324	21,51	1.785	26,06
40 - 49	1.116	20,89	265	17,6	1.381	20,16
50 - 59	817	15,29	227	15,07	1.044	15,24
60 - 69	353	6,61	87	5,78	440	6,42

70 - 79	87	1,63	45	2,99	132	1,93
> 80 anos	14	0,26	11	0,73	25	0,37
Total	5.343	100	1.506	100	6.849	100
Raça/Cor						
Branco	2.319	44,35	610	41,16	2.929	43,64
Pardo	2.025	38,73	578	39	2.603	38,79
Preto	470	8,99	158	10,66	628	9,36
Indígena	17	0,33	16	1,08	33	0,49
Amarelo	35	0,67	14	0,94	49	0,73
Ignorado	363	6,94	106	7,15	469	6,99
Total	5.229	100	1.482	100	6.711	100
Anos de estudo						
Nenhum	49	1	13	0,9	62	1
1 a 4	508	10,4	131	9,4	639	10,2
5 a 8	913	18,7	196	14,1	1109	17,7
9 a 11	1215	24,9	276	19,8	1491	23,8
12 ou mais	861	17,6	341	24,5	1202	19,2
Não aplicável	106	2,2	50	3,6	156	2,5
Ignorado	1.23	25,2	387	27,8	1.617	25,8
Total	4.882	100	1.394	100	6.276	100
Atividades profissionais						
Viajante	931	19,03	303	20,21	1.23	19,31
Turismo	447	9,14	245	16,34	692	10,83
Agricultura	457	9,34	89	5,94	546	8,54
Garimpo	443	9,05	40	2,67	483	7,56
Construtor	239	4,88	3	0,2	242	3,79
Mineração	192	3,92	39	2,6	231	3,61
Doméstico	52	1,06	178	11,87	230	3,6
Caça/Pesca	108	2,21	6	0,4	114	1,78
Pecuária	68	1,39	5	0,33	73	1,14
Exp. vegetal	57	1,16	5	0,33	62	0,97
Outras	1.41	28,88	397	26,48	1.81	28,32
Ignorado	486	9,93	189	12,61	675	10,56
Total	4.89	100	1.5	100	6.39	100

Fonte: Sinan - Ministério da Saúde do Brasil.

Para os casos autóctones, as atividades mais frequentes são “Agricultura” (264 casos; 26,9%) e “Doméstica” (93 casos; 9,5%). Os casos importados são mais frequentes em “Viajantes” (1.110 casos; 22,2%) e “Turismo” (630 casos; 12,5%). As informações preenchidas como “Outras” atividades econômicas representam um total de 25,8% e 28,4% para casos autóctones e importados (Tabela 2). As informações sobre atividades ocupacionais nas notificações de óbitos mostram que os óbitos mais

frequentes na região extra-AMZ são de aposentados (16 óbitos; 13,7%), donas de casa (6 óbitos; 5,1%) e, gerentes em atividades comerciais (6 óbitos; 5,1%).

Entre os casos novos de malária no período, 912 (15%) registros foram notificados como casos autóctones, já importados foram 4.869 (78,9%) e 307 indeterminados (4,5%). Malária por *P. vivax* totalizou 4.592 casos (67,5%) e *P. falciparum* ou casos mistos de malária (*P. falciparum* malária junto com outras espécies) foram responsáveis por 2.159 (31,5%) notificações. Há também um total de 39 casos de malária por *P. malariae* e 59 casos por *P. ovale*, que representam apenas 1,4% do total de casos confirmados ao longo da série histórica (Tabela 2).

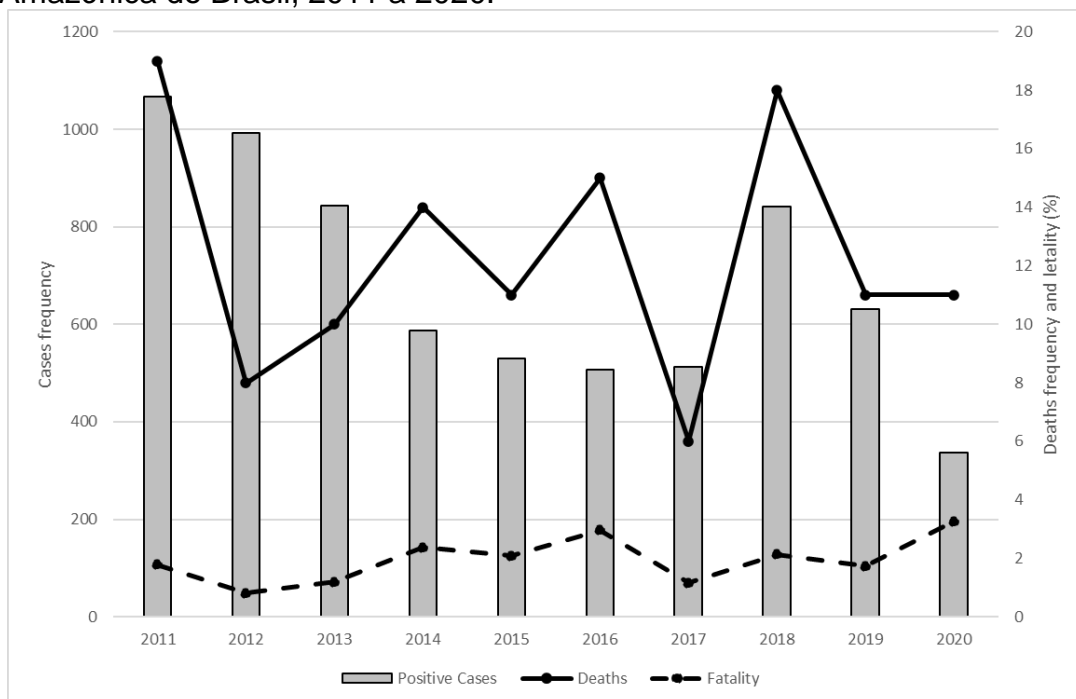
Tabela 2. Distribuição dos casos por espécie de *Plasmodium*, atividade profissional e autoctonia, 2011 a 2020, região extra-AMZ do Brasil.

	Autóctone		Importados		Indeterminados		Total	
	n	%	n	%	n	%	n	%
<b>Tipo de malária para novas infecções</b>								
<i>Plasmodium falciparum</i>	212	21,1	1.805	33,4	96	31,3	2.113	31,4
<i>Plasmodium vivax</i>	785	78	3.539	65,4	200	65,1	4.524	67,3
<i>Plasmodium ovale</i>	2	0,2	42	0,8	5	1,6	49	0,7
<i>Plasmodium malariae</i>	8	0,8	25	0,5	6	2	39	0,6
<b>Atividades económicas</b>								
Viajante	51	5,2	1.111	22,2	62	21,4	1.223	19,5
Turismo	40	4,1	630	12,6	17	5,9	687	10,9
Agricultura	264	26,9	252	5	16	5,5	532	8,5
Garimpo	37	3,8	420	8,4	13	4,5	470	7,5
Construtor	19	1,9	215	4,3	4	1,4	238	3,8
Doméstico	93	9,5	127	2,5	7	2,4	227	3,6
Mineração	25	2,5	180	3,6	21	7,2	226	3,6
Caça/Pesca	15	1,5	95	1,9	3	1	113	1,8
Gado	16	1,6	52	1	2	0,7	70	1,1
Exploração vegetal	22	2,2	40	0,8		-	62	1
Outras	253	25,8	1.423	28,4	96	33,1	1.772	28,2
Ignorado	146	14,9	462	9,2	49	16,9	657	10,5

Fonte: Sinan - Ministério da Saúde do Brasil.

A série temporal apresentou uma média aproximada de 685 casos por ano, sendo o maior número de casos notificados em 2011, com 1.066 casos, reduzindo para 336 casos em 2020. A média mensal foi de 57 casos (mínimo de 10 e máximo de 177). De 2011 a 2020, ocorreram 118 óbitos, com mínimo de cinco óbitos em 2017 e máximo de 18 em 2011 (Figura 6). Destes, 41 (34,7%) óbitos foram por *P. falciparum*, 17 (14,4%) por *P. vivax*, 3 (2,5%) por *P. malariae*, e 56 (47,5%) para malária não especificada. Além das 132 mortes de residentes extra-amazônicos, houve seis óbitos por malária não especificada em residentes da região AMZ que evoluíram para óbito na região extra-AMZ.

Figura 6. Taxa de casos, óbitos e letalidade por malária na região extra-Amazônica do Brasil, 2011 a 2020.



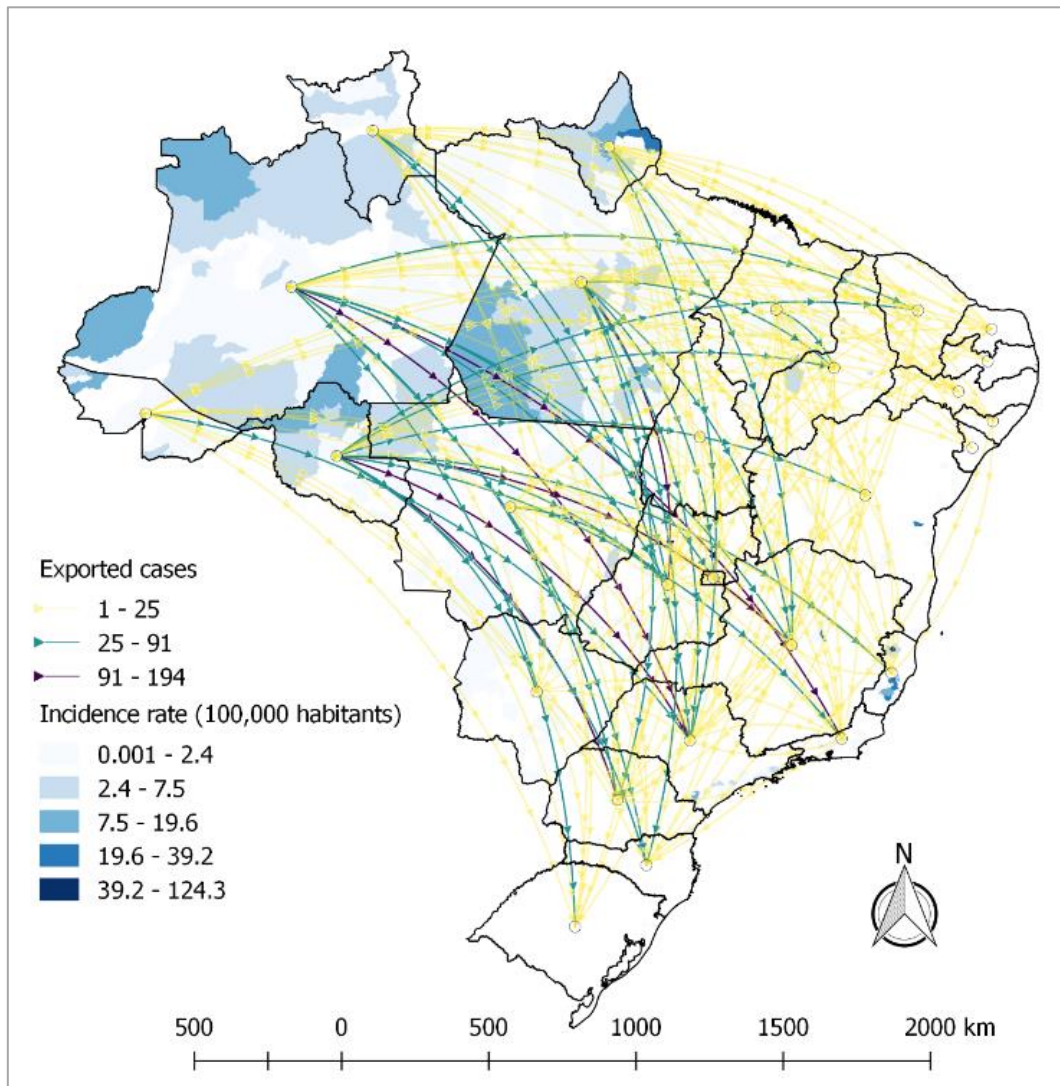
Fonte: Sinan e SIM – Ministério da Saúde do Brasil.

Os casos autóctones mais frequentes de UF foram no Espírito Santo (411 casos; 6%), Bahia (136 casos; 2%) e São Paulo (123 casos; 1,8%). Os estados que mais notificaram casos de malária foram São Paulo (1.405 casos notificados; 20,5%), Rio de Janeiro (778; 11,4%) e Minas Gerais (741; 10,8%). Para os casos importados, a maioria dos casos são notificados em São Paulo

(1.214 casos; 22,4%), Minas Gerais (691 casos; 12,8%) e Rio de Janeiro (649 casos; 12%).

Dos 118 óbitos na região extra-AMZ, a maioria era de residentes de São Paulo (30 óbitos; 25,4%), Minas Gerais (21 óbitos; 17,8%), Bahia e Goiás (9 óbitos cada; 7,6%). Na Figura 7, vemos que os fluxos de casos importados mais intensos foram predominantes na região Norte, principalmente os estados do Amazonas e Rondônia para os estados de São Paulo, Rio de Janeiro e Paraná (Figura 7).

Figura 7. Fluxo de casos importados de malária notificados na região extra-Amazonica do Brasil com taxas de incidência. 2011 a 2020.



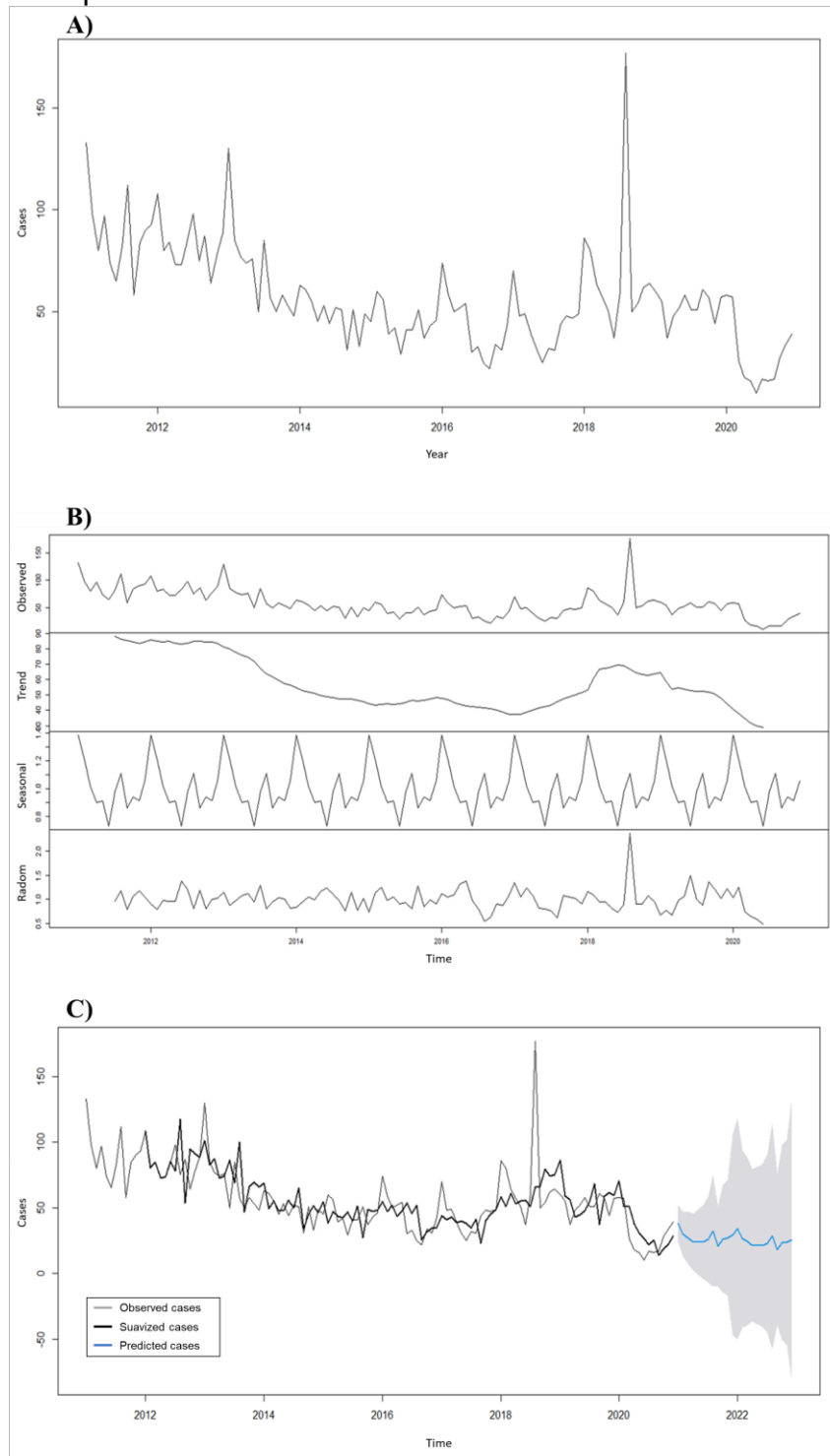
Fonte: Sinan - Ministério da Saúde do Brasil.

As taxas anuais de incidência para toda a região variaram de 0,16 a 0,60 casos por 100.000 habitantes, enquanto as taxas de incidência municipais variaram de 0,008 (São Paulo/SP em 2020) a 1.243 (Vila Pavão/ES em 2018) casos a cada 100.000 habitantes (média: 11; mediana: 4,6). As taxas de notificação apresentaram variação anual semelhante, variando de 0,03 (Salvador/BA em 2020) a 1.286,6 (Vila Pavão/ES em 2018) casos notificados para cada 100.000 habitantes (média 6,7; mediana: 1,6). A taxa de mortalidade por malária extra-AMZ de 2011 a 2020 foi de 1,8, com mínimo de 0,81 em 2012 e máximo de 3,27 em 2020 (Figura 6).

Quanto a outros casos importados, a região notificou 1.771 novas infecções de outros países na série histórica. Os países responsáveis por até 80% do total de casos importados são Angola (53 ; 30,2%), África do Sul (163; 9,2%), Moçambique (158; 8,9%), Guiana Francesa (137; 7,7%), Guiana (132; 7,5%), Nigéria (124; 7,0%), Venezuela (116; 6,5%) e Guiné Equatorial (60; 3,4%).

O modelo de predição utilizado apresentou estimativas para os meses de 2021 e 2022, com média de 27 casos por mês no primeiro ano e 24 casos mensais em 2022. A série temporal apresentou tendência de redução de casos. As estimativas são acompanhadas de intervalos de confiança que, quando previstos para períodos superiores a 12 meses começam a perder precisão em suas estimativas, assim, ressalta-se que para o ano de 2021 estão previstos 324 casos com mínimo de 16 e máximo de 47 casos em um único mês (Figura 8). São esperados 611 casos nos próximos dois anos.

Figura 8. Atributos de série temporal, tendência, sazonalidade e aleatoriedade de casos notificados de malária, por mês, na região extra-Amazônica brasileira com previsão de dois anos.



Fonte: Sinan - Ministério da Saúde do Brasil.

Legenda: A) Séries temporais de malária na região extra- *Amazônica* brasileira; B) Componentes da série temporal da malária; C) Previsão de casos de malária para 2021 e 2022. Em cinza: Intervalos de confiança de 95%.

## DISCUSSÃO

Considerando que o Brasil trabalha com perspectiva de eliminação dos casos de malária até o ano de 2035 [11] analisamos juntos os casos novos de malária e as recorrências no perfil epidemiológico, embora o número de recorrências do banco de dados provavelmente esteja subestimado. Do ponto de vista da eliminação, os casos de recorrência devem ser considerados na análise do perfil, a série temporal, pois um caso de recorrência não tratado pode se tornar fonte de surto, necessitando receber tratamento para malária. Quanto à análise preditiva, utilizamos novos casos de malária apenas porque 79% dos casos notificados na região extra-AMZ foram importados de novas infecções da região AMZ - como no estudo de Machado e colaboradores [18].

Os resultados destacam que a ocorrência de casos de malária em homens, como ocorre na região AMZ, é maior do que em mulheres e que a distribuição dos casos se concentra na idade adulta (20 a 59 anos) e tem distribuição semelhante entre pessoas de cor branca e parda. A distribuição dos casos de malária por faixa etária, para o mesmo período, difere do perfil de casos notificados na região AMZ, onde a concentração de casos em menores de 20 anos é de aproximadamente 46,3% do total de casos. Na região extra-AMZ, a concentração em menores de 20 anos não ultrapassa 9%. Quanto à distribuição dos casos por raça/cor, o perfil também é diferente, pois na região AMZ cerca de 60,8% dos casos são pardos, 17,6% indígenas e apenas 6,2% são brancos (Dados atualizados em 09-06-2021) [19,20].

Considerando que a maioria dos casos notificados na região extra-AMZ são casos importados e que viajar para áreas endêmicas é um conhecido fator de risco para contrair malária [21], essa diferença no perfil pode estar relacionada a questões socioeconômicas, como anos de estudo, atividade econômica e renda. Isso indica que a ocorrência de casos de malária na região extra-AMZ e AMZ do Brasil relaciona-se com as desigualdades sociais.

O MS tradicionalmente apresenta estratégias diferentes para ambas as regiões, porém, sugere-se que sejam trabalhadas metas específicas de controle interno para o país para a região extra-AMZ [11,12,19].

A notificação de casos autóctones na região extra-AMZ é um alerta para os centros de vigilância epidemiológica, pois representa o estabelecimento da transmissão comunitária, que é um risco para a reintrodução da malária nessas regiões não endêmicas, mas receptivas [22]. É importante destacar que aproximadamente 27,5% dos casos extra-amazônicos foram notificados no ano de 2018, ano atípico para o controle da malária no Brasil, pois o país vinha apresentando constantes reduções nas taxas de malária até 2017. Em 2016 o Brasil registrou aproximadamente 129.000 casos e em 2017 relataram mais de 194.000 [11].

Ao analisar o IP e o local de notificação na região extra-AMZ de acordo com a atividade profissional, ficou claro que grande parte dos casos de malária importados são de pessoas com atividades profissionais relacionadas a viagens ou turismo. As atividades de mineração também se destacam. Há uma parcela dos casos classificados como autoctonia indeterminada que são pacientes com atividade profissional também relacionada a viagens [23,24]. As atividades agrícolas se destacam como a atividade mais frequente entre os casos autóctones e isso pode estar relacionado à área em que trabalham ou vivem, considerando que essas atividades podem ser agricultura de subsistência. É aconselhável que estudos específicos para determinar as características da malária relacionada ao trabalho no Brasil sejam realizados diante do fato de que mais de 38% das atividades econômicas estão em branco, ignoradas ou definidas como “Outras”.

Os dados mostraram dois casos autóctones de malária por *P. ovale* no Brasil, que é uma espécie parasitária sem registros anteriores de transmissão comunitária no país [3,22,25]. Essas notificações devem ser analisadas pelas secretarias estaduais e municipais de saúde para melhor compreender esses casos e certificar se houve ou não transmissão comunitária. Essas notificações podem ser erros de preenchimento do sistema ou diagnóstico por se tratar de casos isolados, um em 2011 infectado no estado de Rondônia e notificado no Rio de Janeiro e outro em 2014 infectado e notificado no Espírito Santo.

É importante destacar que 4,4% dos registros do período não tiveram autoctonia definida. Portanto, a investigação epidemiológica extra-AMZ precisa ser aprimorada.

Enquanto a região AMZ apresentar alta incidência de casos autóctones de malária, as notificações de casos extra-amazônicos não diminuirão. É preciso entender que a ocorrência de casos autóctones e importados fora da região AMZ segue um padrão sazonal e uma tendência relacionada ao comportamento da malária na região AMZ [18,23].

É importante que ações e políticas específicas de prevenção de óbitos sejam direcionadas para a região extra-AMZ, principalmente nos estados de São Paulo, Minas Gerais e Goiás, que são os que mais acumulam óbitos nos últimos 10 anos. Esses estados podem ter um número maior de mortes porque outros estados podem não ter infraestrutura para diagnosticar e tratar, levando as pessoas com malária a procurar assistência nesses estados. Pesquisas internacionais destacam alguns fatores de risco comuns para a ocorrência de óbitos, como casos de malária não autóctone, ou provenientes de regiões endêmicas, ou que passam por falhas terapêuticas ou atraso no tratamento [26,27].

Além disso, é necessário avaliar a qualidade dos dados sobre as notificações de óbitos, pois a maior frequência de óbitos foi para a CID-10 B54 que é "Malária não especificada". A literatura [22] sugere uma alta frequência de óbitos por *P. falciparum* na região extra-AMZ, bem como os resultados deste estudo. Recomendamos que sejam exploradas técnicas de articulação entre os sistemas oficiais de informação em saúde para melhor especificação das espécies relacionadas às mortes por malária.

Segundo dados online do Ministério da Saúde do Brasil, as previsões para os meses iniciais de 2021 ficaram próximas do valor registrado no Sinan para casos de malária na região extra-AMZ [20]. Assim, a previsão desenvolvida permite um melhor planejamento de estratégias de controle específicas para a contenção dos casos de malária, principalmente quando se considera o componente de sazonalidade da doença [23,28,29]. A tendência de redução de casos na região extra-AMZ deve ser interpretada

pelo sucesso das ações de controle e prevenção de doenças realizadas pelos programas nacionais e locais. Este modelo pode ser replicado para análises de estado.

A diminuição em 2020 pode estar relacionada com restrições de viagens face à pandemia de COVID-19, reduzindo assim as visitas a zonas endêmicas do Norte do país. Esse comportamento foi observado anteriormente em doenças de notificação compulsória na Austrália e em Taiwan, e isso também pode ter ocorrido no Brasil [30,31].

Quanto às limitações deste estudo, destaca-se o fato de terem sido utilizados dados secundários; a base de dados apresentava algumas lacunas de preenchimento, que apresentavam juntamente com as tabulações dos resultados valores que não coincidem entre si. Essa diferença ocorre porque algumas notificações não possuem todas as informações do formulário de investigação preenchidas. Além disso, a definição de autoctonia de óbito não é possível devido às limitações do banco de dados do SIM. As técnicas de *record linkage* podem ser uma estratégia interessante para apresentar mais informações precisas. Por fim, as atividades ocupacionais no banco de dados do SIM podem não ser precisas; portanto, investigações adicionais são aconselháveis.

## **CONCLUSÕES**

A análise de séries temporais permite visualizar o comportamento epidemiológico da malária frente aos diferentes indicadores de vigilância da malária, acompanhada de análises de previsão de curto prazo que podem servir como ferramentas de apoio à tomada de decisões governamentais.

Para alcançar a eliminação da malária, o governo deve investir na qualificação de pessoal que preencha as notificações de malária e fortalecer o diagnóstico oportuno e adequado da malária, bem como realizar avaliações de qualidade dos bancos de dados do Sinan e SIM.

Assim, um passo para alcançar as metas de eliminação nacional é a eliminação da malária na região extra-AMZ do Brasil, prevenindo casos e mortes pela doença.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. Organização Mundial da Saúde. Relatório mundial da malária 2020: 20 anos de progresso e desafios globais. Genebra: Organização Mundial da Saúde; 2020.
2. Talapko J, Škrlec I, Alebić T, Jukić M, Včev A. Malária: O Passado e o Presente. *Microorganismos*. 2019;7(6):179. Publicado em 21 de junho de 2019. doi:10.3390/microorganisms7060179
3. Oliveira-Ferreira J, Lacerda MV, Brasil P, Ladislau JL, Tauil PL, Daniel-Ribeiro CT. Malária no Brasil: uma visão geral. *Malar J*. 2010; 9:115. Publicado em 30 de abril de 2010. doi:10.1186/1475-2875-9-115
4. Tadei WP, Dutary Thatcher B. Vetores de malária na Amazônia brasileira: *Anopheles* do subgênero Nyssorhynchus. *Rev Inst Med Trop São Paulo*. 2000;42(2):87-94. doi:10.1590/s0036-46652000000200005
5. Hiwat H, Bretas G. Ecologia de *Anopheles darlingi* Root em relação à importância do vetor: uma revisão. *Vetores parasitas*. 2011;4:177. Publicado em 16 de setembro de 2011. doi:10.1186/1756-3305-4-177
6. Sinka ME, Bangs MJ, Manguin S. et al. Um mapa global dos vetores dominantes da malária. *Vetores de parasitas* 5, 69 (2012). <https://doi.org/10.1186/1756-3305-5-69>
7. Duarte EC, Ramalho WM, Tauil PL, Fontes CJ, Pang L. A mudança na distribuição da malária na Amazônia brasileira, 2003-2004 e 2008-2009. *Rev Soc Bras Med Trop*. 2014;47(6):763-769. doi:10.1590/0037-8682-0274-2014
8. França TCC, Santos MG dos, Figueroa-Villar JD. Malária: aspectos históricos e quimioterapia. *Quim Nova*. 2008;31(5):1271–8. Disponível em: <https://doi.org/10.1590/S0100-40422008000500060>

9. Lorenz C, Virginio F, Aguiar BS, Suesdek L, Chiaravalloti-Neto F. Epidemiologia espacial e temporal da malária em regiões extra-AMZs do Brasil. *Malar J.* 2015;14:408. Publicado em 15 de outubro de 2015. doi:10.1186/s12936-015-0934-6.
10. Brasil. Ministério da Saúde do Brasil. Secretaria de Vigilância em Saúde, Departamento de Imunizações e Doenças Transmissíveis, Coordenação Geral de Vigilância de Zoonoses e Doenças Transmissíveis por Vetores. Boletim Epidemiológico: Dia Mundial de Combate à Malária. *Bol Epidemiol* [Internet]. pág. de abril 1-16. 2021. Disponível em: [https://www.gov.br/saude/pt-br/assuntos/media/pdf/2021/abril/26/boletim\\_epidemiologico\\_svs\\_dia-mundial-damalaria.pdf](https://www.gov.br/saude/pt-br/assuntos/media/pdf/2021/abril/26/boletim_epidemiologico_svs_dia-mundial-damalaria.pdf)
11. Brasil. Ministério da Saúde do Brasil. Secretaria de Vigilância em Saúde, Departamento de Imunizações e Doenças Transmissíveis, Coordenação Geral de Vigilância de Zoonoses e Doenças Transmissíveis por Vetores. Boletim Epidemiológico – Malária: 2020 [Internet]. novembro de 2020. p. 118. Disponível em: [https://www.gov.br/saude/ptbr/assuntos/media/pdf/2020/dezembro/03/boletim\\_especial\\_malaria\\_1dez20\\_final.pdf](https://www.gov.br/saude/ptbr/assuntos/media/pdf/2020/dezembro/03/boletim_especial_malaria_1dez20_final.pdf)
12. Brasil. Ministério da Saúde do Brasil. Secretaria de Vigilância em Saúde. Coordenação Geral de Desenvolvimento da Epidemiologia nos Serviços. Guia de Vigilância em Saúde. Volume único Ed. – Brasília: Ministério da Saúde do Brasil, 2019. 740p.
13. Brasil. Guia de tratamento da malária no Brasil. Ministério da Saúde do Brasil, Secretaria de Vigilância em Saúde, Departamento de Imunização e Doenças Transmissíveis. Brasília: Ministério da Saúde do Brasil, 2020. 76 p.: il.
14. Instituto Brasileiro de Geografia e Estatística. Amazônia Legal: O que é [online]. 2021. Disponível em: <https://www.ibge.gov.br/geociencias/cartas-e-mapas/mapasregionais/15819-amazonia-legal.html?=&t=o-que-e>.

15. R Core Team (2019) R: Uma linguagem e ambiente para computação estatística. R Foundation for Statistical Computing, Viena, Áustria. URL <https://www.Rproject.org/>
16. Swapnarekha H, Behera HS, Nayak J, Naik B, Kumar PS. Modelo de inverno multiplicativo de Holts para análise de tendências e previsão da disseminação do COVID-19 na Índia. *SN Comput Sci.* 2021;2(5):416. doi:10.1007/s42979-021-00808-0
17. Koehler AB, Snyder RD, Ord JK. Modelos de previsão e intervalos de previsão para o método multiplicativo de Holt-Winters. *Int J Previsão.* 2001;17(2):269–86. Disponível em: <[https://doi.org/10.1016/S0169-2070\(01\)00081-4](https://doi.org/10.1016/S0169-2070(01)00081-4)>.
18. Machado RLD, Couto ÁAD, Cavasini CE, Calvosa VSP. Malária na região extra- Amazônica: situação no Estado de Santa Catarina. *Rev Soc Bras Med Trop.* 2003;36(5):581–6. Disponível em:<<https://doi.org/10.1590/S0037-86822003000500007>>.
19. Carlos BC, Rona LDP, Christophides GK, Souza-Neto JA. Uma análise abrangente da transmissão da malária no Brasil. *Pathog Glob Saúde.* 2019;113(1):1-13. doi:10.1080/20477724.2019.1581463 Massad E, Laporta GZ, Conn JE, et al. O risco de infecção por malária para viajantes que visitam a região AMZ brasileira: uma abordagem de modelagem matemática. *Travel Med Infect Dis.* 2020;37:101792. doi:10.1016/j.tmaid.2020.101792
20. Brasil. Tableau Public. Dados para Cidadão por Malária - Brasil [ONLINE]. [Internet]. [citado em 9 de junho de 2021]. Disponível em: <https://public.tableau.com/app/profile/mal.ria.brasil>. Acesso em: junho de 2021.
21. de Pina-Costa A, Brasil P, Di Santi SM, et al. Malária no Brasil: o que acontece fora da região endêmica amazônica [correção publicada consta em *Mem Inst Oswaldo Cruz.* 27 de agosto de 2018;113(9):e140228ER]. *Mem Inst Oswaldo Cruz.* 2014;109(5):618-633. doi:10.1590/0074-0276140228
22. Ahmed S, Reithinger R, Kaptoge SK, Ngondi JM. As viagens são um fator de risco chave para a transmissão da malária em cenários de pré-

eliminação na África Subsaariana: uma revisão da literatura e meta-análise. *Am J Trop Med Hyg.* 2020;103(4):1380-1387. doi:10.4269/ajtmh.18-0456

23. Massad E, Laporta GZ, Conn JE, et al. O risco de infecção por malária para viajantes que visitam a região AMZ brasileira: uma abordagem de modelagem matemática. *Travel Med Infect Dis.* 2020;37:101792. doi:10.1016/j.tmaid.2020.101792

24. Massad E, Behrens RH, Burattini MN, et al. Modelagem do risco de malária para viajantes para áreas com transmissão estável de malária. *Malar J* 8, 296 (2009). <https://doi.org/10.1186/1475-2875-8-296>

25. Limongi JE, Costa DC, Carvalho LH, Lopes IS, Silva AA, Ferreira MS (2014) Malária por *Plasmodium ovale* no Brasil: relato de caso importado com período de incubação prolongado. *J Infect Dev Ctries* 8:554-557. doi: 10.3855/jidc.4037

26. Lüthi B, Schlagenhauf P. Fatores de risco associados a mortes por malária em viajantes: uma revisão da literatura. *Travel Med Infect Dis.* 2015;13(1):48-60. doi:10.1016/j.tmaid.2014.04.014

27. Sarkar J, Murhekar MV, Shah NK, et al. Fatores de risco para mortes por malária no distrito de Jalpaiguri, Bengala Ocidental, Índia: evidências para ações futuras. *Malar J* 8, 133 (2009). <https://doi.org/10.1186/1475-2875-8-133>

28. Lechthaler F, Matthys B, Lechthaler-Felber G, et al. Tendências nos casos de malária relatados e os efeitos do controle da malária na República Democrática do Congo. *PLoS One.*2019;14(7):e0219853. Publicado em 2019Jul25. doi:10.1371/journal.pone.0219853

29. Lima MVM, Laporta GZ. Avaliação de modelos de predição da ocorrência de malária no estado do Amapá, Brasil, 1997-2016: um estudo ecológico. Avaliação de modelos de predição para ocorrência de malária no estado do Amapá, 1997-2016: um estudo ecológico. *Epidemiol Serv Saude.* 2021;30(1):e2020080. Publicado em 15 de fevereiro de 2021. doi:10.1590/S1679-49742021000100007

30. Bright A, Glynn-Robinson AJ, Kane S, Wright R, Saul N. O efeito das medidas de saúde pública da COVID-19 em doenças notificáveis nacionalmente na Austrália: análise preliminar. *Commun Dis Intell* (2018).

2020;44:10.33321/cdi.2020.44.85. Publicado em 5 de outubro de 2020.  
doi:10.33321/cdi.2020.44.85

31. Lai CC, Chen SY, Yen MY, Lee PI, Ko WC, Hsueh PR. O impacto da epidemia de doença de coronavírus 2019 em doenças infecciosas de notificação obrigatória em Taiwan: uma análise de banco de dados. *Travel Med Infect Dis.* 2021;40:101997. doi:10.1016/j.tmaid.2021.101997

## **5.2. Descrição do comportamento e cenário epidemiológico da malária na região Amazônica brasileira (Objetivo específico 2)**

### **5.2.1. O Brasil está no caminho de eliminar a malária? Uma análise de séries temporais de casos de malária de 2011 e 2022.**

## **INTRODUÇÃO**

A malária representa uma enfermidade infecciosa aguda e febril prevalente em regiões tropicais. Estimativas globais para o ano de 2021 indicam aproximadamente 247 milhões de indivíduos afetados em todo o mundo [1]. No Brasil, a malária é uma doença de notificação obrigatória para os sistemas de informações epidemiológicas, e a detecção de casos abrange todo o país. A região AMZ responde por 99% dos casos de malária [2]. Os principais agentes causadores são o *Plasmodium vivax* e o *Plasmodium falciparum*, com infecções por *P. vivax* correspondendo a cerca de 83% dos casos na região AMZ [3]. Embora a transmissão de *P. malariae* ocorra, ela é relativamente baixa, enquanto as infecções por *P. ovale* não ocorrem nativamente no Brasil e são relatadas apenas em casos importados de outros países. O principal vetor responsável pela transmissão da malária na região AMZ brasileira é o mosquito *Anopheles darlingi* [2].

Os esforços de controle da malária no Brasil são coordenados pelo Programa Nacional de Prevenção e Controle da Malária (PNCM), estabelecido em 2003 [4], sob a responsabilidade do Ministério da Saúde (MS). O PNCM implementa medidas de controle entomológico e atividades de vigilância epidemiológica em colaboração com as secretarias de saúde estaduais e municipais. O diagnóstico e tratamento da malária estão amplamente disponíveis e são fornecidos gratuitamente em todo o Brasil por meio do Sistema Único de Saúde [2,5]. Em consonância com a com a Estratégia técnica global (ETG) e os Objetivos de Desenvolvimento

Sustentável (ODS), o Brasil visa eliminar a malária do país até 2035, e um plano de eliminação da doença foi publicado em 2022 [6].

O cenário epidemiológico da malária no Brasil ao longo do século XX foi de aumento de casos, principalmente devido ao aumento da urbanização e expansão populacional, juntamente com a exploração crescente da Amazônia na segunda metade do século [7–9]. No início do século XXI, o Brasil registrou mais de 600.000 casos de malária em 2005, mas a incidência tem diminuído desde então, com cerca de 130.000 casos relatados em 2016 [10].

Embora o número de casos de malária tenha diminuído de 2006 a 2022 - com exceção de 2017 e 2018 - houve períodos de aumento no número total, e o número total em 2022 é maior do que em 2016. Esse comportamento compromete a visão de uma estratégia estável de eliminação da malária.

Apesar dos esforços do PNCM para eliminar a doença, mudanças políticas e estruturais no governo do Brasil e no MS enfraqueceram a eficiência do programa desde 2016, quando o PNCM foi fundido com o programa de controle da dengue e perdeu protagonismo e força. Simultaneamente, houve um aumento de casos entre povos indígenas e garimpeiros, principalmente devido à flexibilização das leis de proteção ambiental entre 2019 e 2022 [11]. Além disso, uma análise recente de Laporta et al. (2022) alertou que as ações atuais empreendidas pelo PNCM podem não ser suficientes para alcançar o objetivo de eliminação da malária no país [12].

Portanto, este estudo tem como objetivo analisar o perfil e o cenário epidemiológico dos casos de malária no Brasil. A proposta é obter uma compreensão abrangente das condições prevalentes e como o Brasil pode seguir com as metas de eliminação.

## MÉTODOS

### *Desenho do Estudo*

Trata-se de uma análise de séries temporais para examinar as tendências nas notificações de casos de malária na região AMZ brasileira de 2011 até outubro de 2022. Os dados utilizados foram registros não-nominais obtidos do Sistema de Informação Epidemiológica da Malária no Brasil (Sivep-Malária), fornecidos pelo Ministério da Saúde do Brasil (processos #25072.015038/2021-05, #25072.034317/2021-60 e #25072.004310/2023-85). Os dados populacionais foram obtidos do Instituto Brasileiro de Geografia e Estatística (IBGE).

### *Local do Estudo e População*

O estudo concentrou-se na região AMZ brasileira, que abrange nove estados: Acre, Amazonas, Amapá, Mato Grosso, Maranhão, Pará, Roraima, Rondônia e Tocantins. Estes estados englobam 808 municípios, cobrindo uma área de mais de 5 milhões de km<sup>2</sup>, correspondendo a aproximadamente 58,9% da área total do país. O banco de dados do Sivep-Malária inclui notificações de todos os nove estados, que tiveram uma população estimada de 28,4 milhões em 2022 [13]. A população do estudo compreendeu casos notificados no Sivep-Malária como novos casos de infecção por malária entre janeiro de 2011 e outubro de 2022 (variáveis de filtro: ID\_LVC = 2 e TIPO\_LAM = 1 e 2). Os casos autóctones foram classificados com base em seu local provável de infecção, seguindo as diretrizes fornecidas pelo MS [14, 15].

### *Variáveis do Estudo*

Os casos de malária no Sivep-Malária são notificados como lâminas positivas para a infecção (novas infecção) ou lâminas de verificação de cura (LVC) para possíveis recorrências da doença. O estudo analisou a data das notificações e tratamento, resultados laboratoriais que definem a espécie de *Plasmodium*, sexo, idade, escolaridade, raça/cor, gravidez, presença de sintomas, tipo de detecção, parasitemia, atividades nos últimos 15 dias, local

provável de infecção (país, estado, municípios) e tipo de local provável de infecção.

### *Análise Estatística*

A análise descritiva envolveu a sumarização do número anual de casos de malária. Devido à baixa proporção de malária causada por *P. malariae* e *P. ovale*, esses casos foram analisados juntamente com infecções por *Plasmodium vivax* (*Pv*). Da mesma forma, infecções de espécies mistas que incluíam infecções por *Plasmodium falciparum* foram categorizadas como infecções por *Pf*. As taxas de incidência (IR) foram calculadas dividindo o número total de casos novos pela população estimada de cada município e multiplicando-a por 1.000 residentes, representando a população em risco (IPA - incidência parasitaria anual). As taxas de incidência de malária foram categorizadas com base na estratificação de risco utilizada pelo Ministério da Saúde do Brasil [10], classificando a IR como alto risco (IR  $\geq$  50 casos por 1.000 residentes), médio risco (IR entre 10 e 49,99 casos por 1.000 residentes), baixo risco (IR entre 1 e 9,99 casos por 1.000 residentes) e muito baixo risco (IR  $<$  1 caso por 1.000 residentes).

### *Análises de Tendência*

Descrevemos as tendências dos casos clínicos de malária ao longo do tempo usando o software Joinpoint Regression Versão 4.9.1.0 [16]. A análise utilizou modelos de regressão de Poisson para identificar pontos de inflexão (joinpoints) e Mudanças Percentuais Mensais (MPC) para determinar as tendências (aumento, diminuição ou estabilidade) [17]. Um nível de significância inferior a 0,05 foi considerado para todas as análises estatísticas de tendência. O modelo de regressão joinpoint testa se uma linha de múltiplos segmentos fornece um ajuste estatisticamente melhor para descrever a evolução temporal dos dados em comparação com uma linha reta ou uma linha com menos segmentos [17]. O método de Busca em Grade é empregado para identificar a melhor adequação dos dados para a série temporal, e

permutações de Monte Carlo são usadas para testar e comparar diferentes números de pontos de junção em cada modelo [17,18].

### *Análises de Previsão*

Para prever o comportamento futuro da malária em relação aos objetivos de eliminação da malária do PNCM, foram realizadas análises de previsão no software R Studio (versão 4.1.1.) [19]. Os pacotes *forecast* [20] e *fpp* [21] foram utilizados. O modelo de previsão Holt-Winters (HW) [22,23] foi empregado, considerando aspectos-chave de uma série temporal, como tendência, sazonalidade e aleatoriedade. O método HW é adequado para previsões de curto prazo; portanto, foram usadas contagens mensais de malária autóctone para os próximos 26 meses (anos completos de 2023 e 2024). Para verificar a adequação do modelo, a mesma análise foi realizada para prever casos para 2022, e as observações reais e os valores previstos mostraram resultados semelhantes. O modelo de previsão aplica suavização exponencial e efeitos multiplicativos de sazonalidade. O modelo sazonal multiplicativo HW, que leva em conta a influência de efeitos sazonais em uma doença, foi utilizado [22,23]. A técnica de suavização exponencial do método HW foi empregada para analisar os valores da série temporal e estimar resultados futuros prováveis com um intervalo de confiança de 95%. Os modelos de previsão abrangeram todos os casos autóctones de malária na região AMZ brasileira, bem como análises específicas para infecções por *P. falciparum*, infecções mistas, estados individuais, aldeias de população indígena e áreas de garimpo.

### *Análises Descritivas Espaciais*

O software Qgis foi utilizado para gerar a distribuição de casos de acordo com o município de infecção com base em suas classificações de risco para cada ano. Foram realizadas duas análises espaciais: uma para todos os casos autóctones de malária e outra para casos autóctones de *P. falciparum* e infecções mistas. A análise foi conduzida para descrever o comportamento espacial das taxas de incidência de malária na região AMZ.

### *Considerações Éticas*

Este estudo foi isento de aprovação por comitês de ética, seguindo as diretrizes estabelecidas pela Resolução nº 466/2012 do Conselho Nacional de Saúde do Brasil.

## **RESULTADOS**

### *Epidemiologia da Malária 2011-2022*

Entre janeiro de 2011 e outubro de 2022, a região AMZ registrou um total de 2.034.864 casos de malária. Destes, 279.599 casos (13,7%) foram atribuídos à infecção por *P. falciparum*, enquanto 1.755.265 casos (86,3%) foram causados por *P. vivax*. Ao longo do período de estudo, a proporção de infecções por *P. vivax* variou entre 82,1% (2013) e 89,3% (2018). Em um período de 12 anos, o Brasil conseguiu reduzir mais de 161.000 casos de malária, resultando em uma diminuição de 60,7%. Isso corresponde a uma redução média anual de 14.674 casos (6,4%).

A distribuição de casos de malária em diferentes áreas passou por mudanças significativas. A proporção de casos em áreas rurais diminuiu de 58,0% (154.181) em 2011 para 38,7% (39.266) em 2022. Da mesma forma, a proporção de casos em áreas urbanas caiu de 14,5% (38.485) em 2011 para 8,6% (8.726) em 2022.

Houve um aumento de casos em aldeias de população indígena, de 27.359 em 2011 para 30.398 em 2022. Além disso, as infecções em áreas de garimpo aumentaram de 5,8% (15.332) em 2011 para 16,6% (17.335) em 2022. Também, ocupações relacionadas à agricultura representaram a maioria dos casos (25,4%), e o número de casos entre garimpeiros de ouro teve um aumento de 11,6%, de 5,2% (13.785) em 2011 para 16,7% (17.486) em 2022. A proporção de casos importados de outros países ficou em 2,7%.

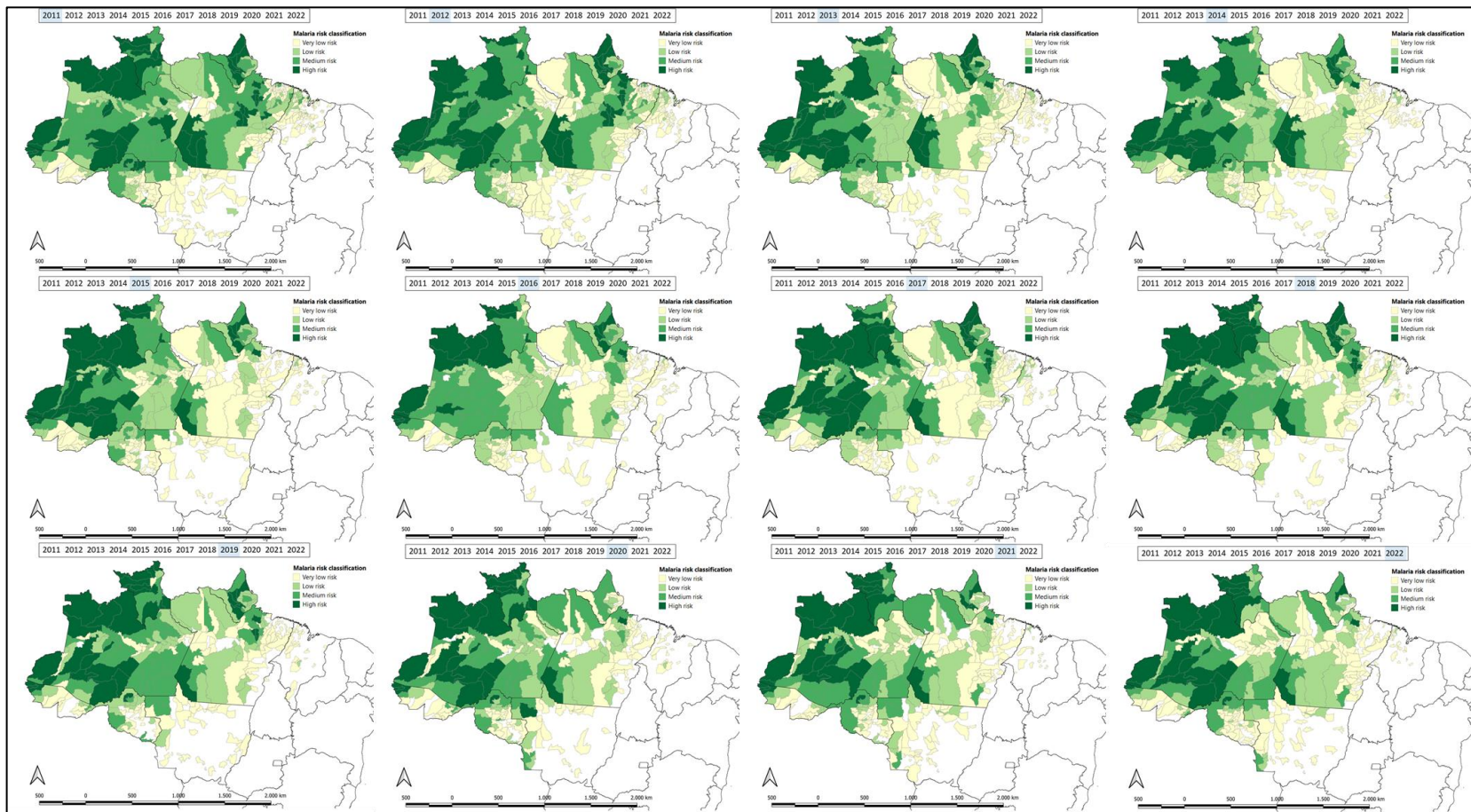
Um total de 384.539 casos de malária foi relatado em aldeias indígenas durante o período de estudo, dos quais 16,5% foram atribuídos a infecções

por *P. falciparum*. Em áreas de garimpo, houve 102.300 casos de malária, com 24,9% causados por *P. falciparum*.

O diagnóstico da malária pode ser feito quando os pacientes procuram atendimento em postos de saúde (detecção passiva) ou quando agentes de saúde buscam casos positivos fora dos postos de saúde (detecção ativa). A taxa média de detecção de casos ativos em aldeias de população indígena correspondeu a 44,7% de todos os casos, com 95,9% dos casos sendo tratados dentro de 24 horas após o início dos sintomas. Em contraste, a taxa média de detecção de casos ativos em áreas de garimpo foi de 10,0%, diminuindo para 5,5% em 2022. A média percentual de casos detectados dentro de 24 horas após o início dos sintomas foi de 33,6%, o que está alinhado com a média geral para a região AMZ. No entanto, a taxa média de detecção oportuna dentro de 24 horas após o início dos sintomas em aldeias indígenas foi maior, atingindo 54,7%.

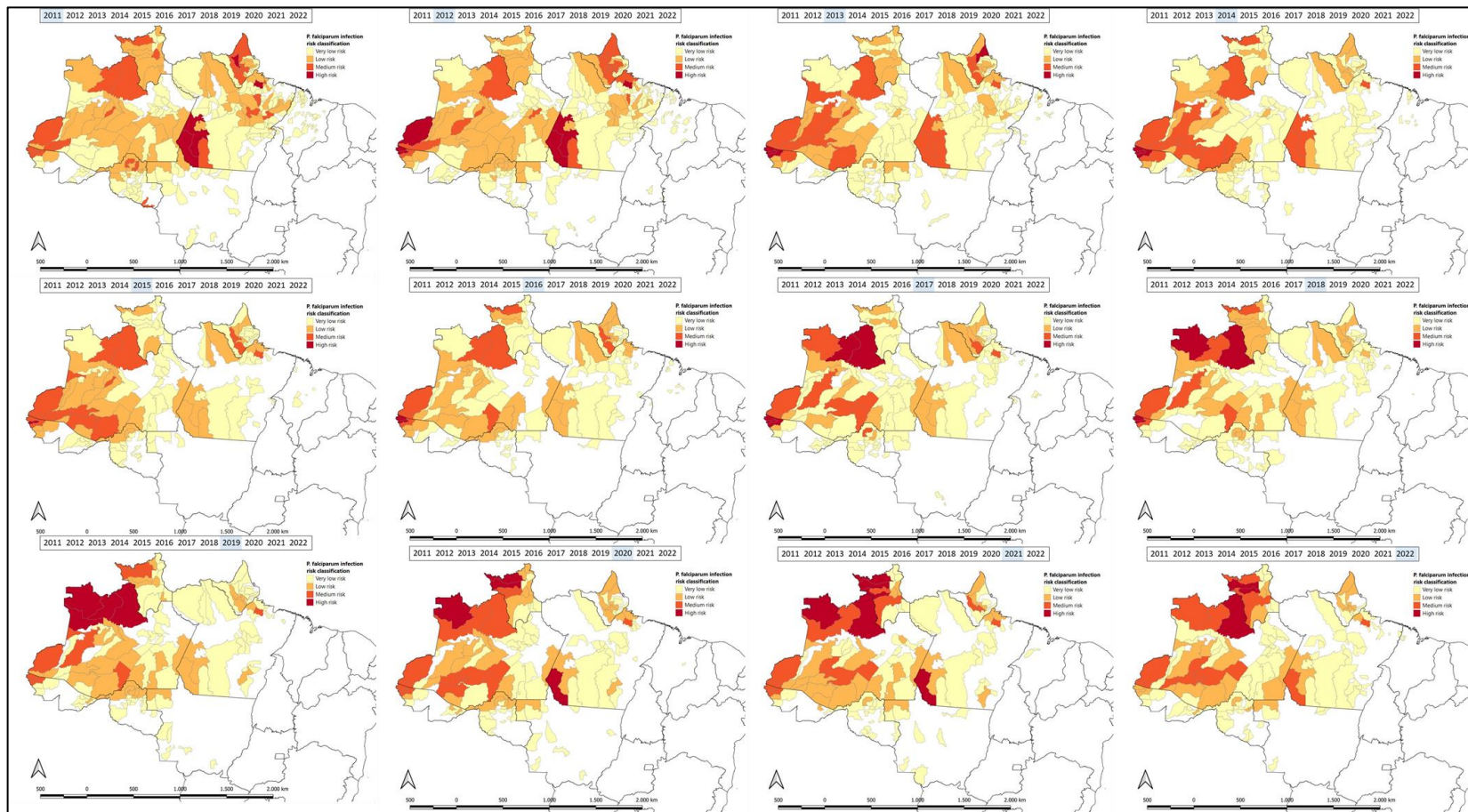
O número de municípios de alto risco (figura 9) diminuiu de 45 em 2011 para 21 em 2022, refletindo uma redução de 46,7%. Da mesma forma, o número total de municípios de médio risco diminuiu de 81 para 39 (uma redução de 51,9%). Embora a redução no número de municípios em relação às classificações de risco para infecções por *P. falciparum* tenha sido mais modesta, houve uma diminuição geral no número de municípios de baixo risco e muito baixo risco. No entanto, nos últimos cinco anos, os municípios de maior risco concentraram-se nos estados do Amazonas e Roraima, principalmente em áreas de mineração e regiões de população indígena (figura 10).

1 Figura 9. Distribuição Espacial de Todas as Classificações de Risco de Infecções por Malária na Amazônia Brasileira, 2021-  
2 2022.



3  
4 Todas as infecções incluídas usando o local de infecção provável. Fonte: Sivep-Malária, Ministério da Saúde do Brasil.

5 Figura 10. Distribuição Espacial de Classificações de Risco de Infecções por *Plasmodium falciparum* na Amazônia Brasileira,  
6 2021-2022.



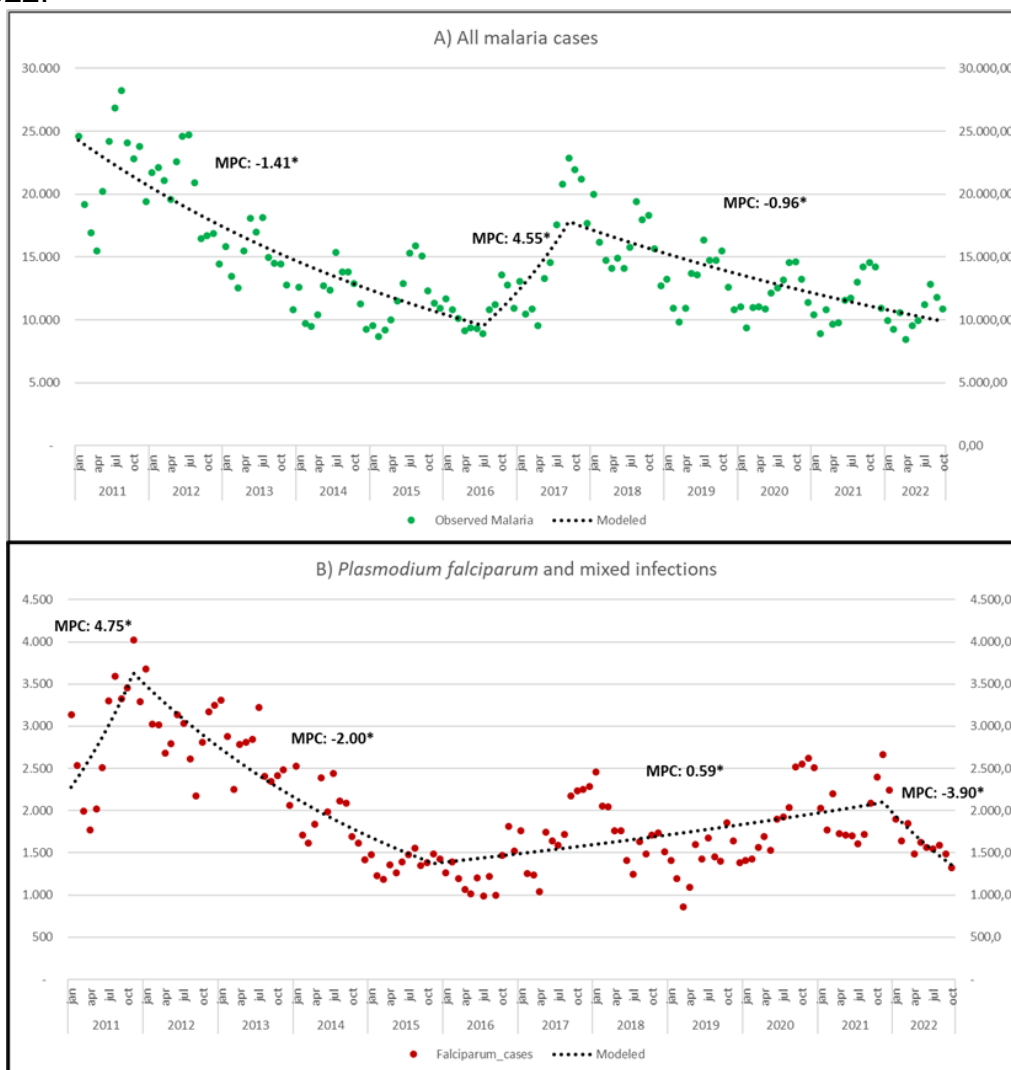
7  
8 \*Infecções por *Plasmodium falciparum* e infecções mistas incluídas usando o local de infecção provável. Fonte: Sivep-  
9 Malária, Ministério da Saúde do Brasil.

### *Análise de Tendência*

A análise de tendência identificou três períodos distintos. O período inicial exibiu uma redução mensal significativa e progressiva de 1,41% até meados de 2016. Posteriormente, houve uma tendência ascendente notável e estatisticamente significativa até a segunda metade de 2017, com um aumento mensal de 4,55%. Essa tendência resultou em um total de mais de 193.000 casos em 2017 e 2018. Após esse período, o Brasil experimentou uma taxa de redução mais lenta (0,96% mensal) em comparação com o período entre 2011 e 2016 (Figura 11A).

Em contraste com o comportamento geral de todos os casos de malária, as infecções por *P. falciparum* exibiram quatro períodos distintos, todos estatisticamente significativos. Houve um aumento em 2011 (aumento mensal de 4,75%), seguido por uma redução mensal sustentada de 2,00% até 2015. De 2016 a 2021, o Brasil testemunhou um aumento gradual, mas consistente, de 0,59% mensal nos casos de *P. falciparum*. Somente no final de 2021 o Brasil exibiu uma tendência substancial de redução quando analisado sem estratificação de ocorrência ocupacional e áreas especiais (redução mensal de 3,90%) (Figura 11B). Quando a análise é feita por algumas características e alguns grupos populacionais, essa tendência não está presente.

Figura 11. Tendências de casos de malária na Amazônia Brasileira, 2011-2022.

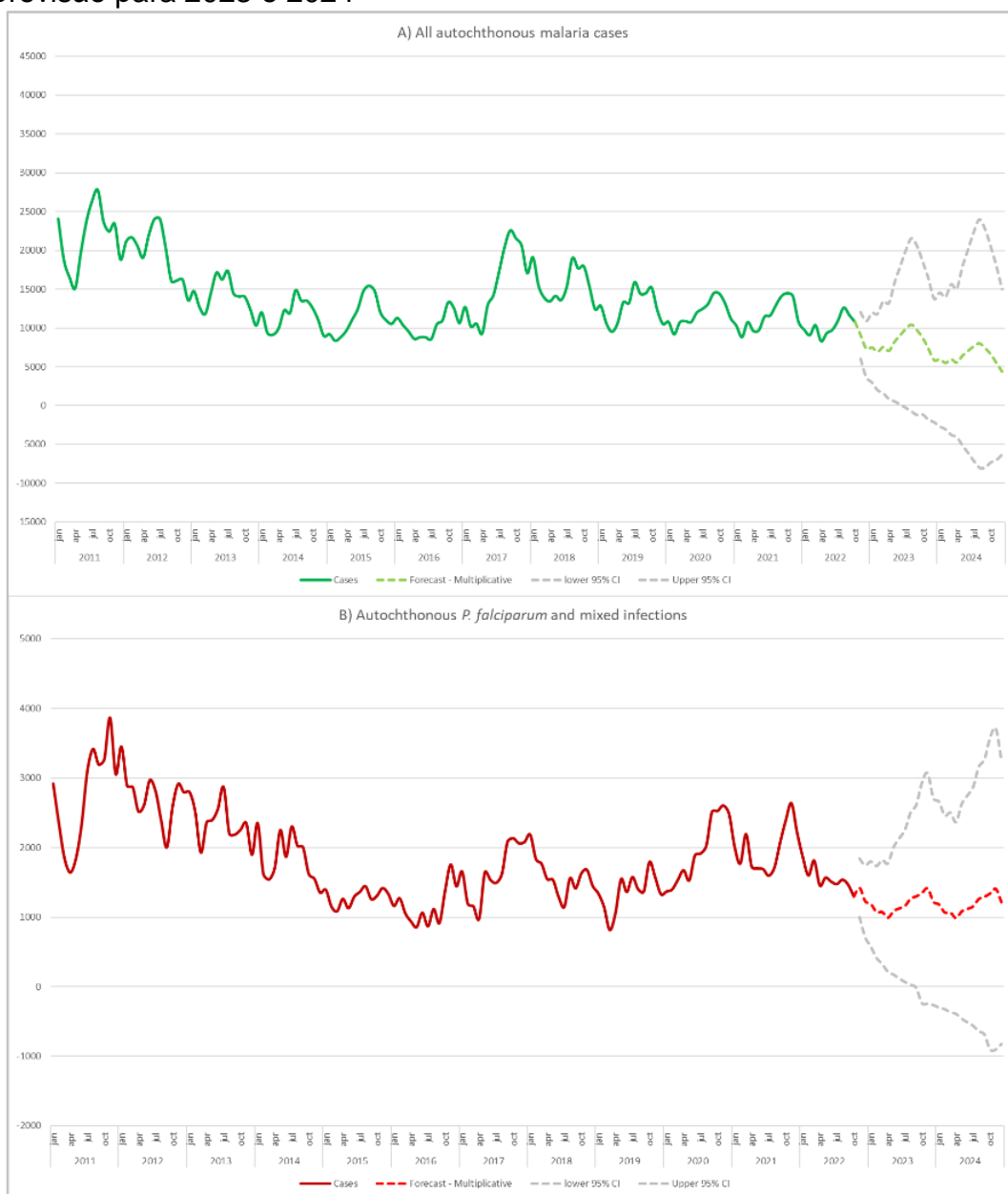


\*\*Legenda: A)\*\* Análise de tendência de todos os casos de malária notificados na região *Amazônica* brasileira entre 2011 e 2022; \*\*B)\*\* Análise de tendência de casos de malária por *P. falciparum* e infecções mistas notificados na região *Amazônica* brasileira entre 2011 e 2022. Fonte: Sivep-Malária, Ministério da Saúde do Brasil.

### Análise de Predição

A previsão para a malária autóctone no Brasil para 2023 e 2024 projetou um total de 97.927 casos de malária em 2023 e 75.796 em 2024 (Figura 12A). O mesmo modelo previu um total de 14.243 casos de *P. falciparum* em 2023 e 14.166 casos em 2024 (Figura 12B).

Figura 12. Todos os casos autóctones de malária na Amazônia Brasileira, previsão para 2023 e 2024



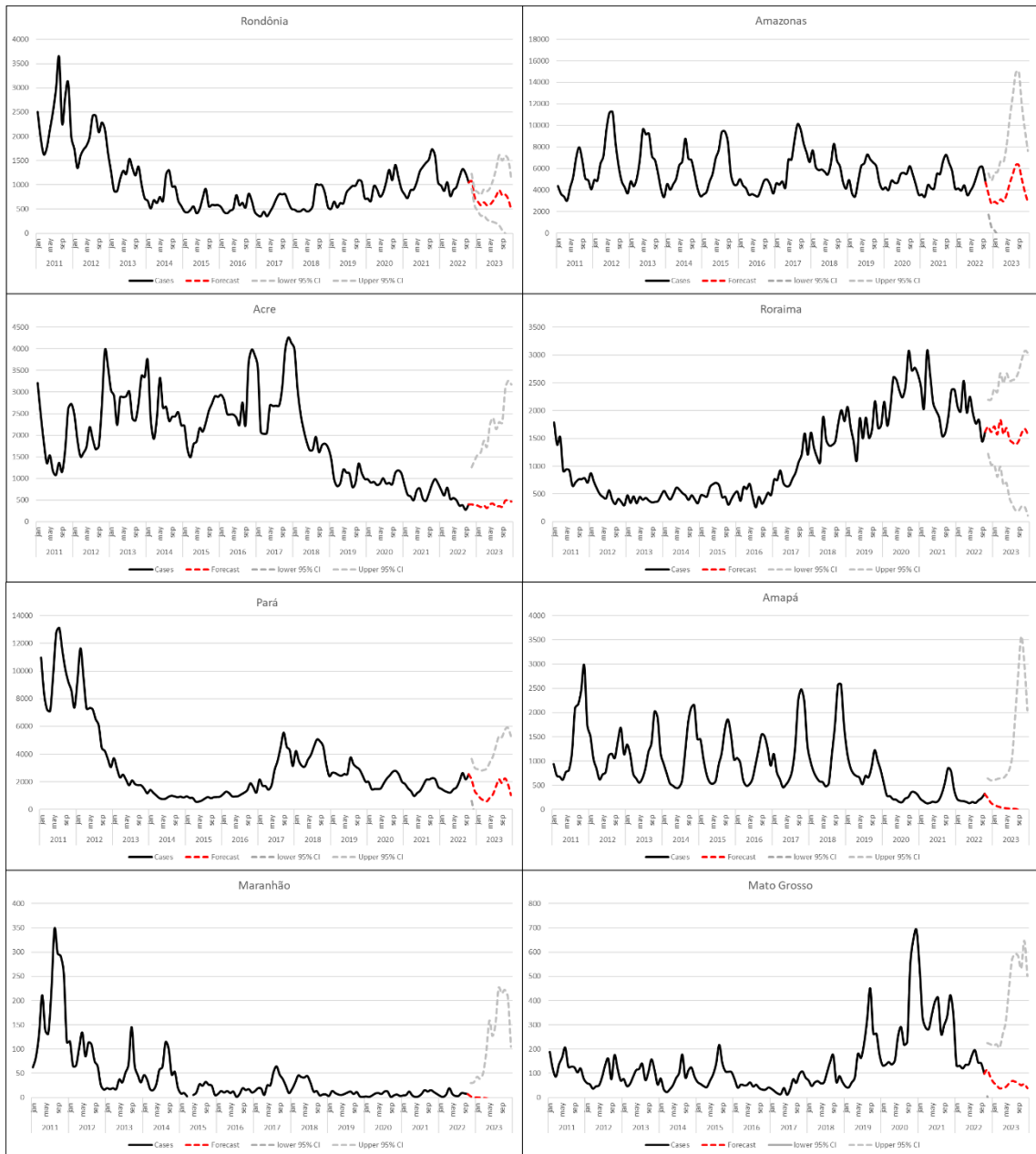
**\*\*Legenda: A)\*\*** Infecções autóctones de malária na Amazônia Brasileira, previsão para 2023 e 2024; **\*\*B)\*\*** Infecções autóctones por Plasmodium falciparum na Amazônia Brasileira, previsão para 2023 e 2024. Fonte: Sivep-Malária, Ministério da Saúde do Brasil.

Considerando que, de acordo com o modelo HW, é concebível que o número de todos os casos autóctones de malária em 2023 possa cair abaixo de 98.000, foram feitas previsões individuais para cada estado endêmico.

Essas previsões revelaram que Acre, Amazonas, Roraima e Pará dificilmente demonstrarão reduções substanciais no número de casos. Pelo contrário, Rondônia, Amapá, Maranhão e Mato Grosso devem apresentar reduções significativas ao longo de 2023 (Figura 13).

Além disso, dado o aumento da ocorrência de casos de malária em áreas indígenas e de garimpo no Brasil, foi crucial prever seu comportamento para 2023. Antecipa-se que os casos de malária nessas regiões continuarão a aumentar, atingindo mais de 37.000 casos em aldeias de população indígena e mais de 21.000 casos em áreas de garimpo em 2023.

Figura 13. Casos autóctones de malária nos estados endêmicos da Amazônia Brasileira, previsão para 2023



Fonte: Sivep-Malária - Ministério da Saúde do Brasil.

## DISCUSSÃO

Entre 2011 e 2022, os casos de malária na região AMZ foram principalmente causados por infecções por *P. vivax*, assim como relatado previamente por Siqueira et al. (2017) [24]. As infecções por *P. vivax* aumentaram, com um aumento observado após 2015. As infecções foram predominantemente observadas em áreas rurais, mas houve um aumento perceptível nas infecções em aldeias de população indígena e áreas de garimpos, especialmente após 2017.

Embora esforços tenham sido feitos para implementar medidas de controle e expandir os serviços de diagnóstico e tratamento entre as populações e territórios indígenas, o controle da malária continua sendo uma preocupação persistente para o PNCM [11]. Isso se deve principalmente à vulnerabilidade social dessas populações (como dificuldades de acesso a serviços de saúde, educação e residência em áreas remotas), o que as torna altamente suscetíveis a cadeias de transmissão em larga escala durante surtos.

O aumento nos casos de malária em aldeias de população indígena, apesar de alguns indicadores epidemiológicos favoráveis, como tratamento oportuno, pode ser atribuído ao aumento do desmatamento na Amazônia observado nos últimos anos [25]. O desmatamento na bacia amazônica tem sido relacionado ao aumento nos casos de malária no Brasil [27,28], e o país também tem testemunhado um aumento nas atividades de garimpo, que estão associadas ao aumento da incidência de malária [25,26]. Dado que muitas áreas de garimpo estão situadas dentro de territórios indígenas [11], é plausível que o desmatamento esteja facilitando a migração de vetores transmissores de doenças da floresta para as aldeias de população indígena. Essa migração aumenta a capacidade vetorial e leva a uma maior ocorrência de casos de malária entre as comunidades indígenas [28–30]. Além disso, um aumento nos casos de *P. falciparum* foi registrado em regiões indígenas e de mineração, o que foi relacionado ao impacto do desmatamento [31].

Para enfrentar essas preocupações, as intervenções de curto prazo em áreas de garimpo podem se concentrar em reforçar o controle vetorial existente e na detecção de casos reativos com esforços de diagnóstico e tratamento rápidos. Embora os indicadores de oportunidades de tratamento para indivíduos infectados em áreas de população indígena sejam favoráveis, há espaço para melhorias na detecção de casos.

O Brasil testemunhou uma redução nos casos de malária na região AMZ, de mais de 264.000 casos anualmente para menos de 140.000 casos em 2021. No entanto, é importante observar que uma avaliação abrangente da incidência de malária no Brasil é insuficiente para tirar conclusões definitivas sobre a eficácia das medidas de controle da malária.

De 2011 a 2016, o Brasil experimentou uma tendência decrescente nos casos de malária, seguida por um aumento nos casos não apenas dentro do país, mas também na região latino-americana em 2017 e 2018 [1]. Posteriormente, o Brasil começou a reduzir consistentemente os casos de malária, exceto as infecções causadas por *Plasmodium falciparum*, que mostraram um aumento contínuo por quase seis anos após inicialmente diminuírem de 2011 a 2015.

Devido ao aumento nas infecções autóctones por *P. falciparum* no início de 2019 e à interrupção das tendências de redução de infecção desde 2016, a previsão para 2023 e 2024 não indica progresso no controle de infecções por *P. falciparum*. O Plano Nacional de Eliminação da Malária previa menos de 6.000 casos de *P. falciparum* em 2023 e 2024 [6]. Enquanto este estudo prevê uma incidência anual de cerca de 14.000 casos. Portanto, ajustes podem ser necessários para as metas estabelecidas pelo PNCM a fim de facilitar uma reavaliação adequada e replanejamento.

A tendência ascendente nas infecções por *P. falciparum*, que persistiu até o final de 2021, sinaliza um enfraquecimento dos esforços de vigilância epidemiológica da malária [11], uma vez que interromper as cadeias de transmissão de *P. falciparum* é crucial [32,33]. Embora os dados atuais sugiram uma tendência de redução, é essencial que os serviços de vigilância

epidemiológica reconheçam o perigo potencial durante os períodos sazonais e adotem planejamento adequado e ações oportunas [34].

Embora pareça que as infecções por *P. falciparum* podem não diminuir nos próximos dois anos, há uma perspectiva positiva para os casos totais de malária em 2023 e 2024, principalmente devido à possibilidade de controle de infecções por *P. vivax*. Se esse cenário se materializar, prevê-se que o PNCM do Brasil reduza para cerca de 75.000 casos até o final de 2024 e menos de 68.000 em 2025. Além disso, a Figura 13 destaca que há a oportunidade de eliminação de *P. falciparum* no Maranhão e Mato Grosso.

O Ministério da Saúde do Brasil deve desenvolver uma política e estratégia abrangente para eliminar a malária no estado do Maranhão. Isso é justificado pelo fato de um número significativo de municípios do estado ter experimentado uma redução nos casos de malária, com uma incidência particularmente baixa de infecções por *P. falciparum* desde 2017.

Embora o estado do Acre tenha uma proporção substancial de municípios de baixo risco, ainda existem áreas de médio risco no oeste e leste. O fortalecimento das estratégias de eliminação nos municípios de baixo risco do Acre pode contribuir para o controle geral da malária no estado. Padrões semelhantes na distribuição espacial das taxas de incidência de malária podem ser observados no estado do Pará, tornando-o adequado para a implementação de estratégias de eliminação em conjunto com o Maranhão, dada sua proximidade geográfica.

Com base na previsão, é projetado que o estado do Amazonas, juntamente com Acre, Roraima e Pará, não experimentará reduções significativas no número de infecções autóctones por malária em 2023. Portanto, esses estados representam desafios significativos nos esforços de controle da malária, como também observado por Laporta et al [12].

Portanto, é recomendável priorizar estratégias de eliminação no Maranhão e nos municípios de baixo risco de Mato Grosso. Além disso, o foco futuro deve ser direcionado para Rondônia, Acre e Amapá para iniciativas de eliminação da malária.

O Plano Nacional de Eliminação da Malária, publicado em 2022 [6], apresentou metas para controlar e eliminar a malária, mas não menciona colaborações com os setores de Vigilância em Saúde Ambiental, Vigilância em Saúde do Trabalhador e Vigilância em Emergências em Saúde Pública do Ministério da Saúde. Entende-se que estes setores têm papel crucial para fortalecer o processo de controle e eliminação da malária no Brasil.

### *Limitações*

As previsões utilizadas neste estudo são adequadas para previsões de curto prazo e não devem ser extrapoladas para cenários distantes. Além disso, a dinâmica dos cenários recentes pode mudar rapidamente devido a intervenções governamentais voltadas para o controle da malária. Portanto, os resultados para o ano de 2024 podem diferir das projeções apresentadas aqui após um ano de medidas estruturadas de controle.

O sistema de informação sobre malária tem sido utilizado para a documentação e vigilância de casos de malária desde sua implementação no início do século [4]. No entanto, subnotificação pode ser maior em aldeias indígenas devido à dificuldade de acesso das equipes de vigilância em saúde e entre regiões de garimpo devido à natureza sensível dessas atividades.

Ainda, apontamos como limitações, os métodos empregados, pois sempre estão alinhados com diferentes graus de incertezas nos seus resultados, não apenas fruto da própria lógica estatística, mas também na construção dos seus pressupostos e algoritmo

## **CONCLUSÕES**

A malária no Brasil está a caminho da eliminação. As análises mostram que há progresso na redução dos casos de malária, porém um ritmo que reduziu na segunda metade da década. Também, eliminação de infecções por *P. vivax* pode ser lenta. No entanto, existem cenários que precisam de atenção, como Amazonas e Roraima. Torna-se evidente que essas áreas enfrentam situações crônicas e críticas. Além disso, o estado atual de

calamidade pública entre as populações indígenas e o aumento dos casos de malária em áreas de garimpo oferecem um risco de perda de controle sobre os casos de malária.

Embora as previsões apresentadas neste artigo indiquem que o Brasil está avançando no controle da malária, o desafio para a eliminação da malária está localizado principalmente nos estados do Amazonas e Roraima, estados que parecem ser fortemente impactados pelo avanço das atividades de garimpo e desmatamento. Áreas multissetoriais para controlar atividades ilegais de garimpo são essenciais para reduzir a malária nesses estados.

Além disso, é importante focar nos estados que estão próximos da eliminação da doença, pois isso contribuirá para a meta do Brasil de eliminar a malária até 2035.

Embora possa não ser suficiente, os próximos quatro anos (2023 - 2026) podem apresentar melhor resultados do que os alcançados entre 2019 e 2022 devido ao incentivo do governo atual para implementar políticas de proteção de populações vulneráveis, promover a saúde de populações indígenas, combater o desmatamento e a garimpo ilegal, e cumprir agendas internacionais, como a Estratégia Técnica Global de malária e os Objetivos de Desenvolvimento Sustentável.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. OMS. World malaria report 2022. 2022.
2. Oliveira-Ferreira J, Lacerda MV, Brasil P, Ladislau JL, Tauil PL, Daniel-Ribeiro CT. Open Access REVIEW Malaria in Brazil: an overview [Internet]. Malaria Journal. 2010. Available from: <http://www.malariajournal.com/content/9/1/115>
3. Siqueira AM, Mesones-Lapouble O, Marchesini P, Sampaio VDS, Brasil P, Tauil PL, et al. *Plasmodium vivax* landscape in Brazil: Scenario and

challenges. *American Journal of Tropical Medicine and Hygiene*. 2016;95:87–96.

4. BRASIL. PORTARIA Nº 1932, DE 09 DE OUTUBRO DE 2003 - Institui o Programa Nacional de Prevenção e Controle da Malária, o Comitê Técnico de Acompanhamento e Assessoramento do Programa e dá outras providências.

5. BRASIL - PNCM. Guia de tratamento da malária no Brasil. 2020. Disponível em: [https://bvsms.saude.gov.br/bvs/publicacoes/guia\\_tratamento\\_malaria\\_brasil.pdf](https://bvsms.saude.gov.br/bvs/publicacoes/guia_tratamento_malaria_brasil.pdf). Acessado em 03 de outubro de 2023.

6. Brasil - PNCM. Elimina Malária Brasil: Plano Nacional de Eliminação da Malária. 1st ed. Vol. 1. Brasília: MS; 2022. 1–60 p. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/svsa/malaria/politicas-de-saude/elimina-malaria-brasil-plano-nacional-de-eliminacao-da-malaria>. Acessado em 03 de outubro de 2023.

7. Bezerra JMT, Barbosa DS, Martins-Melo FR, Werneck GL, Braga ÉM, Tauil PL, et al. Changes in malaria patterns in Brazil over 28 years (1990–2017): results from the Global Burden of Disease Study 2017. *Popul Health Metr*. 2020 Sep 30;18(S1):5.

8. Lima IDSF, Lapouble OMM, Duarte EC. Time trends and changes in the distribution of malaria cases in the Brazilian amazon region, 2004-2013. *Mem Inst Oswaldo Cruz*. 2017 Jan 1;112(1):8–18.

9. Duarte EC, Ramalho WM, Tauil PL, Fontes CJF, Pang L. The changing distribution of malaria in the Brazilian Amazon, 2003-2004 and 2008-2009. *Rev Soc Bras Med Trop*. 2014 Dec;47(6):763–9.

10. Brazil Boletim especial – Malária 2021. Epidemiological Report – Malaria 2021. Disponível em: [https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/especiais/2021/boletim\\_epidemiologico\\_especial\\_malaria\\_2021.pdf/view](https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/especiais/2021/boletim_epidemiologico_especial_malaria_2021.pdf/view). Acessado em: 08 de outubro de 2023.

- 11.. Castro MC, Peterka C. Malaria is increasing in Indigenous and artisanal mining areas in the Brazilian Amazon. *Nat Med*. 2023 Apr 27;29(4):762–4.
12. Laporta GZ, Grillet ME, Rodovalho SR, Massad E, Sallum MAM. Reaching the malaria elimination goal in Brazil: a spatial analysis and time-series study. *Infect Dis Poverty*. 2022 Dec 1;11(1).
13. IBGE. Legal Amazon. <https://www.ibge.gov.br/geociencias/cartas-e-mapas/mapas-regionais/15819-amazonia-legal.html?=&t=o-que-e>.
14. Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Articulação Estratégica de Vigilância em Saúde. Guia de Vigilância em Saúde. 5th ed. Brasilia: MS; 2022. 1–1126 p.
15. OMS. OMS malaria terminology. Geneva, Switzerland: OMS malaria terminology; 2019.
16. Joinpoint. Joinpoint Regression Program.
17. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med*. 2000 Feb 15;19(3):335–51.
18. Bandi P, Silver D, Mijanovich T, Macinko J. Temporal trends in motor vehicle fatalities in the United States, 1968 to 2010 - a joinpoint regression analysis. *Inj Epidemiol*. 2015 Dec 17;2(1):4.
19. R Core Team. R: A language and environment for statistical computing. Vienna, Austria. : URL <https://www.R-project.org/>; 2023.
20. Hyndman RJ, Khandakar Y. Automatic Time Series Forecasting: The forecast Package for R. *J Stat Softw*. 2008;27(3).
21. Hyndman RJ, and AG. Forecasting: principles and practice. Melbourne, Australia: <https://OTexts.org/fpp2/>; 2017.
22. Koehler AB, Snyder RD, Ord JK. Forecasting models and prediction intervals for the multiplicative Holt–Winters method. *Int J Forecast*. 2001 Apr;17(2):269–86.
23. Swapnarekha H, Behera HS, Nayak J, Naik B, Kumar PS. Multiplicative Holts Winter Model for Trend Analysis and Forecasting of COVID-19 Spread in India. *SN Comput Sci*. 2021 Sep 16;2(5):416.

24. Siqueira AM, Bassat Q, Rodovalho S, Lacerda MVG. Raising the red flag for malaria elimination and integrated fever surveillance in the Brazilian amazon. *Lancet Glob Health*. 2017 Mar;5(3):e257–8.
25. Silva-Junior CHL, Silva FB, Arisi BM, Mataveli G, Pessoa ACM, Carvalho NS, et al. Brazilian Amazon indigenous territories under deforestation pressure. *Sci Rep*. 2023 Apr 10;13(1):5851.
26. de Aguiar Barros J, Granja F, Pequeno P, Marchesini P, Ferreira da Cruz M de F. Gold miners augment malaria transmission in indigenous territories of Roraima state, Brazil. *Malar J*. 2022 Nov 29;21(1):358.
27. Wetzler et al. *Malaria Journal* (2022) 21:127 <https://doi.org/10.1186/s12936-022-04141-6> Reserach Changing transmission dynamics among migrant, indigenous and mining populations in a malaria hotspot in Northern Brazil: 2016 to 2020 Erica Anne Wetzler1\*†, Paola Marchesini2†, Leopoldo Villegas3,4 and Sara Canavat Laporta GZ, Ilacqua RC, Bergo ES, Chaves LSM, Rodovalho SR, Moresco GG, et al. Malaria transmission in landscapes with varying deforestation levels and timelines in the Amazon: a longitudinal spatiotemporal study. *Sci Rep*. 2021 Dec 1;11(1).
28. MacDonald AJ, Mordecai EA. Amazon deforestation drives malaria transmission, and malaria burden reduces forest clearing. *Proc Natl Acad Sci U S A*. 2019 Oct 29;116(44):22212–8.
29. Massad E, Coutinho FAB. Vectorial capacity, basic reproduction number, force of infection and all that: formal notation to complete and adjust their classical concepts and equations. *Mem Inst Oswaldo Cruz*. 2012 Jun;107(4):564–7.
30. Sallum MAM, Conn JE, Bergo ES, Laporta GZ, Chaves LSM, Bickersmith SA, et al. Vector competence, vectorial capacity of *Nyssorhynchus darlingi* and the basic reproduction number of *Plasmodium vivax* in agricultural settlements in the Amazonian Region of Brazil. *Malar J*. 2019 Apr 4;18(1).
31. Rerolle F, Dantzer E, Lover A, Marshall JM, Hongvanthong B, Sturrock HJW, et al. Spatio-temporal associations between deforestation and malaria incidence in Lao Pdr. *Elife*. 2021 Mar 1;10.

32. Slater HC, Ross A, Ouédraogo AL, White LJ, Nguon C, Walker PGT, et al. Assessing the impact of next-generation rapid diagnostic tests on *Plasmodium falciparum* malaria elimination strategies. *Nature*. 2015 Dec 2;528(7580):S94–101.
33. Bousema T, Drakeley C. Epidemiology and infectivity of *Plasmodium falciparum* and *Plasmodium vivax* gametocytes in relation to malaria control and elimination. Vol. 24, *Clinical Microbiology Reviews*. 2011. p. 377–410.
34. Reiner RC, Geary M, Atkinson PM, Smith DL, Gething PW. Seasonality of *Plasmodium falciparum* transmission: a systematic review. *Malar J*. 2015 Dec 15;14(1):343.

### **5.3. Descrição dos aspectos metodológicos de *record linkage* para sistemas de informação em saúde de malária (Objetivo específico 3)**

#### **5.3.1. O uso de *record linkage* para recuperação de dados para vigilância de óbitos por malária.**

## **INTRODUÇÃO**

A vigilância em saúde do Brasil possui um rico acervo de dados epidemiológicos represados em diversos Sistemas de Informação em Saúde (SIS). Alguns SIS que podem ser citados como os mais frequentemente utilizados pela Vigilância em Saúde são o Sistema de Informação de Mortalidade (SIM), que contém informações provenientes das Declarações de Óbito (DO); o Sistema de Informação de Agravos de Notificação (Sinan) e Sistema de Informação da Vigilância Epidemiológica (Sivep), em que ambos contém informações provenientes da Fichas de notificação de doenças e agravos de notificação; o Sistema de informações Hospitalares (SIH), que contém informações sobre internações hospitalares; e o Sistema de Informações sobre Nascidos Vivos (SINASC), que contém informações sobre os nascidos vivos do Brasil. [1]

A interoperabilidade destes sistemas ainda não ocorre [1] e não há uma chave única de identificação de paciente (usuário) que permita o relacionamento das diferentes bases de dados, sendo que, historicamente, a maioria dos registros não possuem o número de identificação do cadastro de pessoa física (CPF) ou registro geral (RG) dos usuários. Isso faz com que, caso haja necessidade de relacionamento de informações em diferentes bases, seja necessária a vinculação de dados destes SIS por técnicas probabilísticas de relacionamento de dados. Essas técnicas utilizam medidas

de similaridade entre variáveis identificadoras dos sujeitos como nome, nome da mãe, data de nascimento, sexo, local de residência, entre outras. [2]

O relacionamento probabilístico de dados tem apresentado bons resultados em pesquisas de diferentes áreas do conhecimento e fornecido evidências em saúde de qualidade para os setores da Vigilância em saúde [3-5]. Porém, a vinculação, além de depender de recursos humanos especializados, depende também da qualidade dos dados presentes nestes SIS e do poder de processamento computacional disponível para essa tarefa. [6]

Com a necessidade de integrar informações provenientes de diferentes SIS e frente a limitações relacionadas a qualidade e acesso aos dados, os métodos de vinculação probabilística de dados (*record linkage*) tem se mostrado como uma alternativa viável para esta integração. [7] Esse método tem se mostrado versátil, pois trabalha com um conjunto de regras que permite analisar a similaridade entre os registros. A técnica permite ainda analisar os possíveis pares relacionados como verdadeiros ou não e maximizar o número de pares verdadeiros identificados. [8]

Os SIS utilizados na vigilância epidemiológica da malária são o Sivep-Malária e o Sinan. Casos de malária notificados na região AMZ (estados do Norte mais os estados do Maranhão e Mato Grosso) são notificados no Sivep-Malária e casos da região extra-AMZ são notificados no Sinan. Estas fichas de notificação individual não coletam informações sobre a evolução do caso ou mortalidade. [9,10]

Para o monitoramento oficial de óbitos por malária pela vigilância epidemiológica da malária do Programa Nacional de Controle e Prevenção da Malária (PNCM) são utilizados os dados do SIM, que abarcam óbitos causados pelas diferentes espécies de malária. Para se investigar questões específicas da vigilância de óbitos por malária, é necessário que se analise as informações dos diferentes SIS em conjunto.

É importante que se tenha clareza sobre o que se está relacionado à ocorrência de óbitos por malária, principalmente ao se considerar que este é um evento de saúde evitável. [11] Com estas informações é possível promover

ações de prevenção a estes óbitos. Então, frente a necessidade de qualificar ações da vigilância do óbito por malária, este estudo considerou importante a operacionalização desse relacionamento, o relato metodológico do processo e a descrição das informações recuperadas entre os diferentes SIS. Para que assim se possa discutir os procedimentos realizados, os desafios encontrados e ações que podem ser desenvolvidas com os novos dados disponíveis.

O objetivo deste estudo é descrever os resultados e os processos metodológicos utilizados na vinculação de dados (*record linkage*) de óbitos por malária e casos da doença. Espera-se que este material possa servir como modelo para outros processos de *record linkage*.

## **MÉTODOS**

Estudo transversal descritivo, relatando processos metodológicos de técnica de *record linkage* para validação da técnica para recuperação de dados sociodemográficos e epidemiológicos de óbitos por malária.

São descritos procedimentos e resultados de um método de vinculação de dados de três Sistemas de Informação em Saúde. Neste trabalho foram descritas as variáveis sociodemográficas e epidemiológicas presentes no Sistema de Informação de Mortalidade (SIM), do Sistema de Vigilância Epidemiológica da Malária (Sivep-Malária) e do Sistema de Informação de Agravos de Notificação (Sinan). Os dados são referentes ao total de óbitos por malária declarados no SIM (CID-10: B50, B51, B52, B53 e B54) e ao total de casos positivos de malária notificados no Brasil, no Sivep-Malária e Sinan, entre 2011 e 2020.

Do banco de dados do SIM foram utilizadas as seguintes variáveis: data do óbito, data de nascimento, nome do paciente, nome da mãe, sexo, idade, raça/cor, escolaridade, causa básica do óbito, causas associadas ao óbito, local de ocorrência do óbito, assistência médica, investigação do óbito e ocupação.

Do banco de dados do Sivep-Malária e Sinan foram utilizadas as seguintes variáveis: data de início de sintomas, data de notificação, data de nascimento, nome do paciente, nome da mãe, sexo, idade, raça/cor, escolaridade, resultado do exame laboratorial e tipo de lâmina (se lâmina de verificação de cura ou não).

Para descrição dos dados se utilizou técnicas de estatística descritiva. Foram calculados os intervalos de confiança para as frequências relativas das variáveis estudadas para averiguação de semelhança entre as distribuições dos dados pareados pela técnica de *record linkage* em comparação aos dados originais. O teste qui-quadrado foi aplicado para verificar diferenças entre os registros originais e os dados vinculados. Foi adotado um nível de significância de 0,05. Quando o valor de p for maior que 0,05 significa que não há diferença entre as proporções das variáveis entre os bancos. Ou seja, os registros vinculados, nesse cenário, apresentarão distribuição das variáveis de forma similar ao do banco original. Os registros dos SIS também foram comparados para identificar possíveis inconsistências entre as variáveis sociodemográficas e as informações de espécie de plasmódio. Também se verificou a quantidade de informações marcadas como “ignoradas”, “em branco” ou informações mal definidas como “malária não especificada” que foram recuperadas entre os bancos.

O acesso aos dados ocorreu por meio de solicitação, via ofício, ao gabinete da Secretaria de Vigilância em Saúde e Ambiente (SVSA) do Ministério da Saúde (MS), de modo que a autorização de manuseio destes dados foi concedida pelas áreas da Coordenação-Geral de Informações e Análises Epidemiológicas (CGIAE) e Coordenação-Geral de Vigilância de Zoonoses e Doenças de Transmissão Vetorial (CGZV). O manuseio foi realizado em ambiente seguro e monitorado pela CGIAE, seguindo protocolos de segurança para evitar a má utilização ou vazamento dos dados.

Este trabalho compõe o projeto de pesquisa intitulado “Distribuição espaço-temporal de óbitos por malária com estudo de seus fatores de risco e causas associadas no Brasil de 2011 a 2020”, que recebeu parecer favorável

de nº 51246121.0.0000.5558 pelo Comitê de Ética em Pesquisa da Faculdade de Medicina da Universidade de Brasília.

Para desenvolvimento do *record linkage* se utilizou a interface R Studio com técnicas de vinculação probabilística, descritas adiante e a linguagem R (versão 4.2.1). A vinculação probabilística pode ser entendida como uma técnica que considera um conjunto de regras de combinação que analisam a probabilidade de registros distintos pertencerem ao mesmo indivíduo por meio do cálculo de medidas de similaridade entre variáveis de campo aberto, categóricas, datas ou medidas de dissimilaridade entre variáveis numéricas. [7]

### *Procedimentos específicos da vinculação*

#### *Planejamento prévio*

Foram definidas as variáveis de interesse dos bancos de dados do SIM, Sivep-Malária e Sinan. Para o pareamento, selecionou-se como variáveis chaves de vinculação “Nome do paciente”, “Data de nascimento” e “Nome da mãe”, seguindo o que foi sugerido em outros estudos. [6,12-13]

#### *Preparação dos dados*

Para realizar o tratamento dos dados e as vinculações foram utilizados os pacotes *tidyverse*, *lubridate*, *stringr*, *stringi*, *fuzzyjoin*, *stringdist*, *abjutils* e *Rcpp* do software R.

Para a preparação dos dados foram usados como modelo os métodos de limpeza e tratamento de dados já relatados por Garcia et al., (2022) em pesquisa prévia.[6] Assim, para diminuir a quantidade de erros e inconsistência uma padronização inicial das variáveis chaves foi realizada nos três bancos de dados. Todos os caracteres das variáveis “nome do paciente” e “nome da mãe” foram transformados em caracteres maiúsculos. Os conectores “ E ”, “ A ”, “ DA ”, “ DE ”, “ DO ”, “ DAS ”, “ DES ” e “ DOS ”, os acentos gráficos, os símbolos e os espaços em branco foram retirados. Ainda, textos com caráter de “observações” também foram retirados do campo “Nome do paciente”, como: números, pontuações, “KG” (quilo), “PESO”, “TEL”

ou “TL” ou “FONE” (telefone), “OBS” (observação) e as palavras “ignorado” ou “ignorada”.

A variável “nome do paciente” foi separada em duas novas variáveis: “primeiro nome” e “sobrenome” a fim de aumentar a sensibilidade do método de detecção de pares. Todos os procedimentos descritos foram aplicados nas bases de dados do SIM, Sivep-Malária e Sinan.

Devido ao elevado número de notificações do banco de dados Sivep-Malária, as notificações deste sistema foram separadas em blocos bianuais (2011-2012, 2012-2013 ... 2018-2019, 2019-2020), para facilitar o processamento dos dados pelo software e ainda captar informações epidemiológicas prévias ao óbito. Considerando que o adoecimento e morte por malária é um evento agudo [14] entende-se não ter havido perda de pares neste processo de blocagem.

#### *Parâmetros e regras para a vinculação*

Para a realização da vinculação probabilística utilizou-se o pacote “fuzzyjoin”, que para o seu funcionamento é necessário definir parâmetros de comparação das informações dos diferentes bancos de dados. Nesse pacote, é necessário definir medidas de distância para as variáveis do tipo nominais e numéricas, além de datas. Duas medidas de distância utilizadas nos pareamentos são a de Levenshtein [15] e a de Cosine, [16] a primeira usada para nomes e a segunda para sobrenomes. Para datas, pode-se optar por utilizar estas mesmas técnicas de distâncias, considerando um possível preenchimento incorreto.

Os valores definidos como parâmetros são os valores padrão, considerados arbitrariamente razoáveis em estudos de vinculação probabilística. [12] Assim, para a distância de Levenshtein, foi parametrizado o valor 3 e para Cosine o valor 0,5.

#### *Execução da vinculação*

Após a definição dos parâmetros e das regras de vinculação, realizou-se a vinculação do tipo *inner join*, em que são selecionados os pares que

aparecem simultaneamente nos conjuntos de dados de dois sistemas de informação, por meio da função “fuzzy\_inner\_join”, do pacote “fuzzyjoin”, [17] da linguagem R (versão 4.1.1). Os diferentes tipos de *linkage* podem ser verificados no relato metodológico feito por Garcia e colaboradores (2022). [6]

A vinculação considerou, de forma simultânea, as seguintes chaves: “primeiro nome”, “sobrenomes”, “data de nascimento” e “nome da mãe”. Para “primeiro nome” e “data de nascimento” utilizou-se a distância de Levenshtein e para “sobrenomes” e “nome da mãe” utilizou-se a distância de Cosine. Foram vinculados os dados do Sivep-Malária e do Sinan aos dados do SIM. O resultado de cada vinculação foi unificado para a construção de um único banco de dados. Considerando que os dados do Sinan e Sivep-Malária podem apresentar mais de uma notificação para a mesma pessoa, o software R entende que a notificação do banco principal (SIM) pode encontrar mais de um par, e assim o software pode apresentar mais de um resultado pareado para a mesma notificação do SIM. Nestas situações, optou-se por utilizar as notificações com informações epidemiológicas mais recentes, com base na data de início de sintomas dos bancos do Sivep-Malária ou Sinan.

#### *Regras de exclusão de falsos-positivos*

Considerando o número elevado de registros pós-vinculação, foi inviável a realização da revisão manual dos pares, assim como foi relatado por Grannis et al. [18]. Desse modo, estabeleceu-se uma regra de validação que considera como verdadeiros-positivos apenas os pares em que o conteúdo de pelo menos uma das três variáveis-chaves, utilizadas no pareamento, fossem idênticas entre os diferentes sistemas de informação. Após esta etapa fez-se a revisão manual dos pares e nenhum par falso-positivo foi encontrado.

#### *Análise dos resultados*

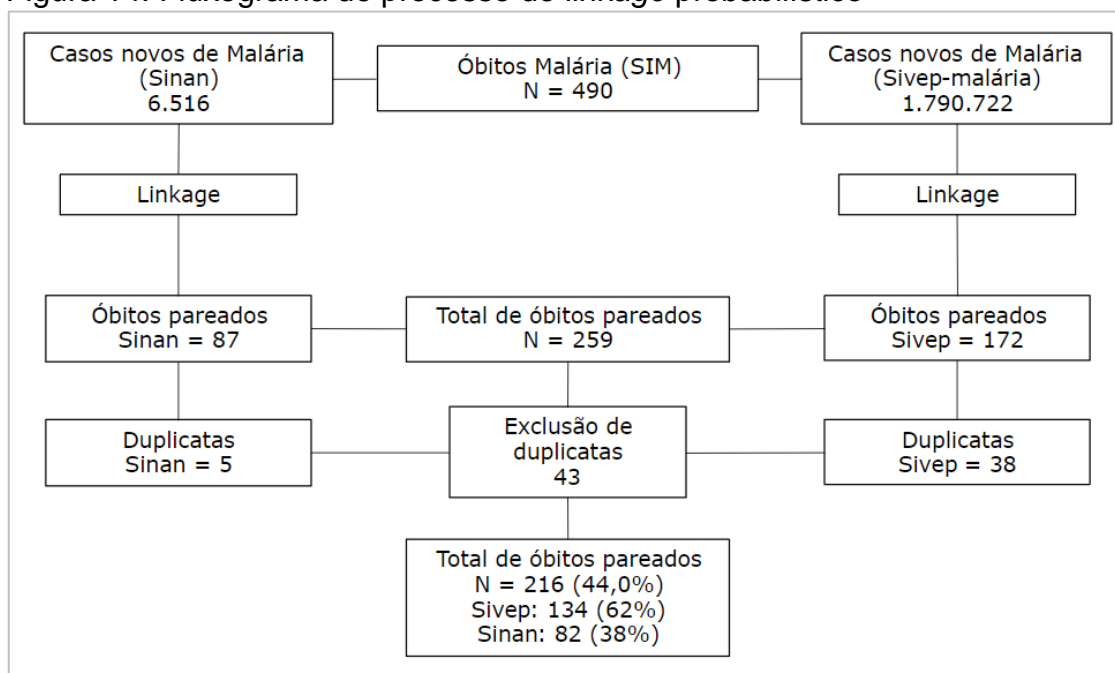
Os resultados foram analisados por meio de estatística descritiva. O banco vinculado foi comparado ao banco original do SIM a fim de verificar se a distribuição das proporções do banco vinculado era similar às

proporções das variáveis do banco original do SIM. Para isto calculou-se os intervalos de confiança das proporções das distribuições das variáveis do banco vinculado e aplicou-se o teste de qui-quadrado para verificar se houve diferença entre os grupos estatisticamente significativa.

## **RESULTADOS**

Os dados do SIM totalizaram 490 óbitos entre 2011 e 2020, o Sivep-Malária apresentou 1.790.722 notificações e o Sinan somou 6.516 registros. Após a realização da vinculação SIM-Sinan e a exclusão dos registros considerados falsos-positivos, pelo critério estabelecido no estudo, foram encontrados 87 pares. Já a vinculação SIM-Sivep resultou em 172 pares. Ao todo foram encontrados 259 pares. Dos 259 pares foram identificadas 43 duplicatas (38 na vinculação SIM-Sivep e 5 na vinculação SIM-Sinan). Restaram 134 óbitos da vinculação SIM-Sivep e 82 da SIM-Sinan, totalizando 216 óbitos com informações recuperadas, equivalente a 44,0% dos 490 óbitos registrado no período de estudo (Figura 14).

Figura 14. Fluxograma do processo de *linkage* probabilístico



Fonte: Autoria própria.

Dos 216 óbitos vinculados, 148 (68,5%) apresentaram “nome do paciente” com concordância perfeita entre os SIS, 117 (54,2%) apresentaram o “nome da mãe” com concordância perfeita e 166 (76,9%) registros apresentaram concordância perfeita para a variável “data de nascimento”. Quanto ao resultado da combinação das variáveis chaves, 65 (30,1%) apresentaram uma equivalência perfeita (nome do paciente, nome da mãe e data de nascimento idênticos), 85 (39,4%) apresentaram pareamento com duas variáveis idênticas e 66 (30,5%) apresentaram apenas 1 variável idêntica (Tabela 3).

Tabela 3. Quantidade de pareamentos por chaves-de-vinculação utilizadas.

Concordância de:	Variáveis utilizadas	Pares	%
3 variáveis	Nome do paciente	65	30,1%
	Nome da mãe		
	Data de nascimento		
2 variáveis	Nome do paciente	22	10,2%
	Nome da mãe		
	Nome do paciente	43	19,9%
	Data de nascimento		
1 variável	Nome da mãe	20	9,3%
	Data de nascimento		
	Nome do paciente	18	8,3%
	Nome da mãe		
	Data de nascimento	38	17,6%
	<b>Total</b>	<b>216</b>	<b>100,0%</b>

Fonte: Autoria própria.

Os anos com maior número de óbitos pareados fora 2012 com 32 pares (48,4% dos óbitos registrados no ano), e o ano com menor quantidade de pares encontrados foi 2015, com apenas 12 pares (31,5% dos óbitos no ano). Em média, foram registrados 49 óbitos por ano, dos quais aproximadamente 21 foram pareados, ou seja, uma média de 44,0% dos óbitos anuais tivera complementação de informações epidemiológicas a partir dos dados do Sivep-Malária e Sinan.

A verificação de semelhança entre as distribuições dos dados pareados (N = 216) apontou que as proporções das variáveis “sexo”, “raça/cor”, “causa básica do óbito”, “assistência médica”, “investigação do óbito”, “ano de ocorrência do óbito” e “local de ocorrência do óbito” apresentaram, de uma forma geral, distribuições similares às do banco original (N = 490). A distribuição das variáveis “Faixa etária” e “Escolaridade” foram as únicas em que as distribuições não foram similares às do banco original (Tabela 4).

Apesar dos dados, de uma forma geral terem apresentado distribuições semelhantes, algumas variáveis não tiveram distribuição semelhante entre o banco vinculado e os bancos de dados originais: se esperava 14,3% de óbitos entre 31 e 40 e foram pareados 21,3% (IC 95%: 16% - 27%), isso indica que houve um pareamento de mais pares entre 31 e 40 anos que o que se

esperava. O inverso ocorreu entre as faixas etárias de 71 a 80 anos e maiores de 80 anos que tiveram um pareamento menor que o esperado (Tabela 4).

Tabela 4. Perfil da mortalidade por malária e dos óbitos pareados entre 2011 e 2020, Brasil.

	Total de óbitos N = 490	%	Óbitos pareados N = 216	%	IC 95%	P-valor
<b>Sexo</b>						0.8
Masculino	285	58,2	128	59,3	(52 - 66)	
Feminino	205	41,8	88	40,7	(34 - 48)	
<b>Faixa etária</b>						0.014*
< 1	24	4,9	9	4,2	(2,0 - 8,0)	
1 - 10	40	8,2	21	9,7	(6,3 - 15)	
11 - 20	38	7,8	21	9,7	(6,3 - 15)	
21 - 30	56	11,4	29	13,4	(9,3 - 19)	
31 - 40	70	14,3	46	21,3	(16 - 27)	
41 - 50	54	11	32	14,8	(10 - 20)	
51 - 60	63	12,9	25	11,6	(7,8 - 17)	
61 - 70	49	10	14	6,5	(3,7 - 11)	
71 - 80	47	9,6	9	4,2	(2,0 - 8,0)	
> 80	48	9,8	10	4,6	(2,4 - 8,6)	
Em branco	1	0,2	-	-	-	
<b>Raça/cor</b>						0.8
Branca	120	24,5	53	24,5	(19 - 31)	
Preta	35	7,1	14	6,5	(3,7 - 11)	
Amarela	2	0,4	-	-	-	
Parda	258	52,7	112	51,9	(45 - 59)	
Indígena	65	13,3	29	13,4	(9,3 - 19)	
Em branco	10	2	8	3,7	(1,7 - 7,4)	
<b>Anos de estudo</b>						0.004*
Nenhuma;	86	17,6	18	8,3	(5,2 - 13)	
1 a 3 anos;	70	14,3	24	11,1	(7,4 - 16)	
4 a 7 anos;	74	15,1	32	14,8	(10 - 20)	
8 a 11 anos;	76	15,5	44	20,4	(15 - 26)	
12 anos ou mais;	56	11,4	42	19,4	(15 - 25)	
Ignorado.	48	9,8	19	8,8	(5,5 - 14)	
Em branco	80	16,3	37	17,1	(12 - 23)	
<b>Causa básica do óbito (CID-10)</b>						0.2
B50 - <i>Plasmodium falciparum</i> (Pf)	101	20,6	54	25	(19 - 31)	
B51 - <i>Plasmodium vivax</i> (Pv)	183	37,3	79	36,6	(30 - 43)	
B52 - <i>Plasmodium malariae</i> (Pm)	5	1	3	1,4	(0,36 - 4,3)	
B53 - Outras Formas	6	1,2	5	2,3	(0,86 - 5,6)	
B54 - Não Especificada	151	30,8	66	30,6	(25 - 37)	
Malária - causa associada	44	9	9	4,2	(2,0 - 8,0)	
<b>Teve assistência médica?</b>						0.4
Sim	311	63,5	150	69,4	(63 - 75)	
Não	49	10	17	7,9	(4,8 - 13)	
Ignorado	21	4,3	10	4,6	(2,4 - 8,6)	
Em branco	109	22,2	39	18,1	(13 - 24)	
<b>O óbito foi investigado?</b>						0.3
Sim	94	19,2	51	23,6	(18 - 30)	
Não	236	48,2	94	43,5	(37 - 50)	
Em branco	160	32,7	71	32,9	(27 - 40)	

<b>Ano de ocorrência</b>						>0.9
2011	72	14,7	29	13,4	(9,3 - 19)	
2012	66	13,5	32	14,8	(10 - 20)	
2013	45	9,2	19	8,8	(5,5 - 14)	
2014	38	7,8	20	9,3	(5,9 - 14)	
2015	38	7,8	12	5,6	(3,0 - 9,7)	
2016	39	8	19	8,8	(5,5 - 14)	
2017	39	8	15	6,9	(4,1 - 11)	
2018	61	12,4	30	13,9	(9,7 - 19)	
2019	39	8	14	6,5	(3,7 - 11)	
2020	53	10,8	26	12	(8,2 - 17)	
<b>UF de ocorrência do óbito</b>						0.9
Rondônia	37	7,6	14	6,5	(3,7 - 11)	
Acre	22	4,5	3	1,4	(0,36 - 4,3)	
Amazonas	97	19,8	40	18,5	(14 - 24)	
Roraima	51	10,4	17	7,9	(4,8 - 13)	
Pará	73	14,9	28	13	(8,9 - 18)	
Amapá	31	6,3	18	8,3	(5,2 - 13)	
Tocantins	2	0,4	1	0,5	(0,02 - 3,0)	
Maranhão	20	4,1	4	1,9	(0,59 - 5,0)	
Piauí	6	1,2	4	1,9	(0,59 - 5,0)	
Ceará	3	0,6	1	0,5	(0,02 - 3,0)	
Rio Grande do Norte	3	0,6	1	0,5	(0,02 - 3,0)	
Paraíba	2	0,4	-	-	-	
Pernambuco	7	1,4	2	0,9	(0,16 - 3,7)	
Alagoas	1	0,2	1	0,5	(0,02 - 3,0)	
Sergipe	1	0,2	1	0,5	(0,02 - 3,0)	
Bahia	11	2,2	6	2,8	(1,1 - 6,2)	
Minas Gerais	25	5,1	13	6	(3,4 - 10)	
Espírito Santo	5	1	3	1,4	(0,36 - 4,3)	
Rio de Janeiro	9	1,8	7	3,2	(1,4 - 6,8)	
São Paulo	33	6,7	23	10,6	(7,0 - 16)	
Paraná	5	1	3	1,4	(0,36 - 4,3)	
Santa Catarina	6	1,2	3	1,4	(0,36 - 4,3)	
Rio Grande do Sul	5	1	3	1,4	(0,36 - 4,3)	
Mato Grosso do Sul	2	0,4	1	0,5	(0,02 - 3,0)	
Mato Grosso	16	3,3	8	3,7	(1,7 - 7,4)	
Goiás	10	2	5	2,3	(0,86 - 5,6)	
Distrito Federal	7	1,4	6	2,8	(1,1 - 6,2)	

Fonte: Sivep-Malaria, SIM. \*P-valor < 0,05

Também não se encontraram distribuições semelhantes para pessoas sem escolaridade, foram pareados 8,3% (IC 95%: 5,2% - 13,0%) dos óbitos e o esperado era 17,6%; pessoas com mais de 11 anos de estudo, foram pareadas 19,4% (IC 95%: 15,0% - 25,0%) e o esperado era 11,4%.

Apesar de o teste qui-quadrado apontar que não houve diferença entre os grupos, ao se analisar a variável “Causa do óbito” observa-se que houve menor pareamento para óbitos por malária como causa associada (4,2%; IC 95%: 2,0% - 8,0%) que o esperado (9,0%). Isso foi visto também na variável

“UF de ocorrência” onde foram pareados menos óbitos (1,4%; IC 95%: 0,3% - 4,3%) que o esperado (4,5%) no estado do Acre e em São Paulo encontrou-se mais óbitos pareados (10,6%;) que o esperado (6,7%).

As causas de óbito descritas no SIM apresentaram alguns dados consistentes, e outros não, ao se comparar com os dados recuperados dos SIS Sivep-Malária e Sinan. Os óbitos por *Pf* (N=54; CID-10: B50) tiveram 20% (N=11) de inconsistência com os dados epidemiológicos, estes apontavam infecções por *Pv* ao invés de *Pf*. Os óbitos por *Pv* (N=79; B51) tiveram apenas 5,0% (N=4) de inconsistências detectadas. Dos três óbitos por *Pm* (B52), dois foram marcados como infecções por *Pf*. Os cinco óbitos por "outras formas de malária" (B53) apresentaram testes positivos para *Pf* (N=1), *Pv* (N=3) e infecções mistas (N=1). Já os dados recuperados para óbitos marcados como "Malária não especificada" (N=66) foram de 39% (N=26) infecções por *Pf*, 48% (N=32) infecções *Pv*, 7,6% (N=5) infecções mistas, 1 (1,5%) por *Pm*, 1 (1,5%) por *Po* e 1 (1,5%) resultado para "não *falciparum*". Os óbitos onde malária não foi a causa básica do óbito (N=9) foram 7 (78%) infecções por *Pv* 1 (11%) por *Pf* e 1 (11%) infecção mista (Tabela 5).

Tabela 5. Causas básicas de óbito (SIM) e recuperação de resultados laboratoriais (Sivep-Malária e Sinan).

	B50 – <i>Pf</i> . N = 54	B51 - <i>Pv</i> , N = 79	B52 – <i>Pm</i> N = 3	B53 - Outras formas. N = 5	B54 - Não especificada. N = 66	Causa associada. N = 9	Total
<i>Pf</i>	39 (72,2%)	2 (2,5%)	2 (66,7%)	1 (20,0%)	26 (39,4%)	1 (11,1%)	71 (32,9%)
<i>Pv</i>	10 (18,5%)	75 (94,9%)	0 (0,0%)	3 (60,0%)	32 (48,5%)	7 (77,8%)	127 (58,8%)
Mista	4 (7,4%)	1 (1,3%)	0 (0,0%)	1 (20,0%)	5 (7,6%)	1 (11,1%)	12 (5,6%)
<i>Pm</i>	0 (0,0%)	0 (0,0%)	1 (33,3%)	0 (0,0%)	1 (1,5%)	0 (0,0%)	2 (0,9%)
<i>P. ovale</i>	1 (1,9%)	0 (0,0%)	0 (0,0%)	0 (0,0%)	1 (1,5%)	0 (0,0%)	2 (0,9%)
Não-F	0 (0,0%)	1 (1,3%)	0 (0,0%)	0 (0,0%)	1 (1,5%)	0 (0,0%)	2 (0,9%)
<b>Total</b>	54 (25,0%)	79 (36,6%)	3 (1,4%)	5 (2,3%)	66 (30,6%)	9 (4,2%)	216 (100,0%)

Fonte: Sivep-Malaria, SIM.

Quanto à recuperação de dados para preenchimentos “ignorados” ou “em branco” para as demais variáveis dos SIS, para o campo “escolaridade” recuperou-se 37 (75,5%) registros dos 49 marcados como “ignorado” ou “em branco” nos SIS; para “raça/cor” foram recuperados 8 (72,7%) de 11 registros

“ignorado” ou “em branco” no Sivep-Malária ou Sinan; para o campo “ocupação” no SIM e “atividade nos últimos 15 dias” foram recuperados 56 registros (60,2%) de ocupação no SIM em relação a 93 preenchimentos marcados como “ignorados” ou “em branco” na variável de “atividades nos últimos 15 dias” no Sivep-Malária ou Sinan.

Para a análise de inconsistências não se considerou os preenchimentos “ignorado” ou “em branco” das variáveis. Dos 216 registros apenas 3 (1,4%) apresentaram inconsistências para a variável “Sexo”; para “nível de escolaridade” houve uma consistência de 25,9% entre o preenchimento do SIM e dos outros SIS; para “raça/cor” houve uma consistência de 86,1% entre os sistemas; para “faixa etária” a consistência foi de 95,3%.

## **DISCUSSÃO**

Com uma estrutura de vigilância em saúde sendo constituída de diversos SIS é importante que processos de vinculação probabilística entre bases de dados sejam realizados, principalmente devido à ausência de interoperabilidade entre estes sistemas e a ausência de uma chave de identificação única para os usuários do SUS nos sistemas de informação em saúde.

Em 2018, foi publicada a Lei Geral de Proteção de Dados (LGPD) (Lei nº 13.709, de 14 de agosto de 2018) [19] que preza pelo acesso e tratamento de dados pessoais. A publicação da lei trouxe novas regras de acesso a informações e regras mais rígidas para o acesso de dados nominais para fins de pesquisa. A partir da LGPD, os dados pessoais são de domínio do titular (pessoa física), que podem negar a disposição do dado para fins de pesquisa. Além disso, para que dados pessoais sejam disponibilizados, é necessário que o titular autorize previamente sua disponibilização para qualquer finalidade, por meio de um termo de consentimento.

Por isso, a LGPD preconiza que os dados disponibilizados para fins de pesquisa sejam anonimizados sempre que possível. Para pesquisas de saúde

pública, a LGPD permite a disponibilização de dados pessoais, desde que tratados exclusivamente dentro do órgão e estritamente para a finalidade de realização de estudos. Esses novos processos podem tornar o processo de execução da pesquisa mais longo e demorado que o inicialmente programado. Para o meio acadêmico, a LGPD tem gerado certa mora e prejuízo a ciência nas universidades.

As vinculações de dados têm permitido a identificação de novas informações para a vigilância em saúde e isso subsidia processos de avaliação de qualidade de dados, [20] complementação de informações [21] e construção de políticas públicas em saúde. As informações permitem conhecer o perfil de usuários na área da saúde, permitem a realização de análises individuais, acompanhamento de coortes e compreensão dos caminhos e itinerários percorridos por populações nos serviços de saúde. [22]

Assim, um motivador para a realização deste estudo foi o objetivo de consolidar uma colaboração entre o ambiente acadêmico e o Programa Nacional de Prevenção e Controle da Malária (PNCM). Pretende-se que os dados encontrados neste processo de vinculação possam ser utilizados também pelos serviços de saúde, para criação de políticas, identificação de riscos e planejamento de ações de controle e prevenção de óbitos.

O estudo apresentou uma recuperação de 44,0% do total de óbitos. Isso indica que há baixa qualidade de preenchimento de informações pessoais, em ambos os sistemas. Anteriormente, a má qualidade de informações já foi relatada como um limitador em processos de vinculações probabilísticas. [8] Isso desperta um alerta que além de capacitar os núcleos e técnicos da vigilância epidemiológica também é necessário fazer esta atividade junto aos profissionais de saúde que fazem os registros das declarações de óbito.

Notou-se um equilíbrio entre as distribuições de casos entre as variáveis, isto indica que a técnica de *record linkage* utilizada é capaz de parear registros com distribuições de variáveis semelhantes à do banco original.

É visível na Tabela 3 que a utilização de métodos que só permitem o pareamento de registros onde ocorre apenas a perfeita concordância entre os registros tem um aproveitamento de apenas 30,1% do total de óbitos que podem ser pareados. Isso evidencia que a utilização de métodos probabilísticos para vinculação de dados incrementa pelo menos duas vezes mais o número de pares vinculados.

As chaves utilizadas mostraram que há uma maior frequência de registros com o "data de nascimento" preenchido corretamente entre as bases, isso pode indicar que em outros bancos de dados esse cenário seja similar.

Ainda, ao se olhar os pares que possuíram apenas uma variável chave idêntica entre os SIS, a "data de nascimento" foi a que permitiu o maior número de pareamentos quando as informações de outras duas variáveis chaves não eram similares (38 pares; 17,6%) entre os sistemas.

Nesse sentido, considerando os dois pontos anteriores, indica-se fortemente que a variável "data de nascimento" seja frequentemente utilizada em processos de vinculação probabilística, pois ela tende a apresentar menor quantidade de erros de preenchimento comparado a variável "nome do paciente" e "nome da mãe".

Como as regras de vinculação consideram possíveis erros de grafia no preenchimento das variáveis-chaves (nome, nome da mãe e data de nascimento) para o pareamento, registros de óbitos podem não ter sido pareados pelo excesso de erros ortográficos no preenchimento das fichas do Sivep-Malária e Sinan. Então, o baixo pareamento de registros de pessoas acima de 71 anos pode estar relacionado com a baixa escolaridade do usuário. Em contrapartida, entre usuários de 31 a 40 anos, que tiveram maior pareamento, a escolaridade predominante é de acima de 8 anos de estudo. Ele se vê entre óbitos registrado com a causa básica "Malária como causa associada", pois a maioria destes registros é de usuários com "Nenhuma" escolaridade.

O pareamento maior ou menor que o esperado entre algumas variáveis está relacionado ao nível de escolaridade da pessoa que evoluiu a óbito.

Dessa forma, além de destacar que a qualidade do dado é um ponto crítico para otimizar o processo de vinculação de dados entre diferentes SIS, também se destaca a importância que os técnicos da vigilância epidemiológica deem maior atenção quando estiverem coletando informação de pessoas que possuam menor nível de escolaridade ou que estejam em idades mais avançadas (idosos).

Quanto às informações recuperadas, cerca de 30,6% das informações a respeito das espécies de malária dos óbitos não foram especificadas na DO, contudo o *record linkage* permitiu a identificação destas informações, estimando que quase a metade dos óbitos por malária não especificada são causadas por infecções de *Plasmodium vivax* e quase 40% são óbitos são por infecções de *Plasmodium falciparum*.

A vigilância do óbito por malária - a nível nacional, estadual e municipal - pode utilizar estas informações para direcionar ações de controle da doença para prevenir óbitos, uma vez que com o *record linkage*, além de se conhecer a causa do óbito pode se detalhar estas causas básicas previamente inespecíficas, e assim investigar os locais de óbitos (registrados no SIM) e os locais de prováveis infecções (registrados no Sivep-Malária e Sinan) com maior precisão de informações.

A união das informações destes SIS permite que os serviços de vigilância ampliem e aprimorem a investigação de óbitos, permitindo a identificação de situações de saúde mais específica. No caso, a estimativa inicial era que cerca de 25,0% dos óbitos pareados eram por *Plasmodium falciparum* e 36,6% por *Plasmodium vivax*, com a recuperação de dados estas estimativas passam a ser de 37% e 51,4%, respectivamente. Isso evidencia uma fragilidade na classificação da causa de óbito no SIM e um alerta de que os óbitos por *Plasmodium falciparum* podem estar sendo subestimados, em quase 50%, no Brasil.

O *record linkage* também permitiu recuperação de dados sociodemográficos referentes ao nível de escolaridade, raça/cor e ocupações/atividades econômicas dos pacientes. Essas informações podem ser utilizadas pelos serviços de vigilância epidemiológica para refinar as

análises de perfis de mortalidade e assim direcionar as ações de prevenções de óbitos com maior precisão.

Em contramão a este refinamento de informações, é preciso que se investigue as inconsistências de dados que também foram verificadas. É visível que a consistência das informações sobre sexo, faixa etária e raça/cor foram altas (acima de 92%) por serem informações básicas de preenchimento obrigatório nas fichas e as informações sobre escolaridade apresentaram mais da metade dos registros pareados com inconsistências entre os dados do SIM com os do Sivep-Malária ou Sinan. Destaca-se novamente que esta é uma informação de elevada importância para complementar a definição do perfil de mortalidade e as análises de situação em saúde para aprimorar ações voltadas para grupos vulneráveis.

Dessa forma, é importante que os setores da vigilância epidemiológica investiguem e entendam qual o motivo destas inconsistências, pois podem estar relacionadas à forma como o dado é coletado durante a notificação do caso ou durante o preenchimento da declaração de óbito, e esta investigação pode trazer à luz outras dificuldades e limitações no processo de coleta de dados.

Além disso, é preciso investigar os registros de óbitos no SIM que não foram pareados, por ser uma informação sensível é algo que não pôde ser realizado neste trabalho. Então, sugere-se que o PNCM possa verificar esta questão para identificar óbitos que possivelmente não foram notificados no Sivep-Malária ou Sinan.

Assim, como os SIS da malária não apresentam informações específicas sobre óbitos pela doença, o mesmo ocorre sobre o registro de internações pela doença. O monitoramento e a investigação de fatores associados às hospitalizações por malária podem ajudar a prevenir estas situações, que conseqüentemente podem impactar na redução de óbitos. Sendo assim, recomenda-se que este processo de vinculação de dados também seja realizado entre o Sistema de Informações Hospitalares, Sivep-Malária e Sinan. Além disso, recomenda-se a incorporação de um campo “evolução” na ficha de notificação individual do Sivep-malária e Sinan seja

avaliada no PNCM, isto permitiria o acompanhamento do paciente e coletar informações sobre cura, recorrências da doença, transferências (entre unidades de saúde), abandono do tratamento, internações ou óbito.

Outro problema é a inexistência de uma variável identificadora única nas diferentes bases de dados. Atualmente, os SIS já são elaborados com modelagens de dados que utilizam documentos de identificação, como o *CPF* ou o número no cadastro do Cartão SUS, como variáveis de registro obrigatório nos SIS. Um exemplo disso é o e-SUS Notifica, [23] lançado em 2020.

A utilização destas técnicas e realização deste trabalho são motivados devido à falta de interoperabilidade entre os diferentes Sistemas de Informação em Saúde do Ministério da Saúde. Porém, os cenários epidemiológicos atuais urgem pela integração destas diferentes informações, vide a dificuldade de se trabalhar com dados de diferentes fontes.

É importante que o departamento de informática de responsabilidade destes sistemas em conjunto com os departamentos de saúde, construam projetos e/ou atualizações dos SIS de forma a tornarem estes dados relacionáveis e, dentro da anonimidade que se preconiza para os dados sensíveis, permitir a identificação de dados do mesmo usuário, em diferentes dos SIS, nos dados públicos que são disponibilizados pelos entes federados que permitam análises mais qualificadas da situação de saúde da população.

Quanto aos métodos empregados neste estudo, é importante destacar que a definição das regras e dos parâmetros para a vinculação podem ser ajustadas conforme o nível da qualidade dos dados que serão utilizados. A utilização das distâncias de Levenshtein [15] e a de Cosine [16] podem ser calibradas e assim aumentar o número de pares encontrados, porém este aumento pode fazer necessária a revisão manual dos pares para exclusão de falsos-positivos e a depender da magnitude dos resultados esta revisão manual pode se tornar inviável. [18]

Dessa forma, alternativas podem ser investigadas, como por exemplo a utilização de métodos de vinculação determinística [6] ou outros pacotes e

funções disponíveis para o software R como o *fastlink*, [24] ou até mesmo fazer a utilização de outros softwares como Reclink [25] ou Python. [26]

Por fim, este estudo apresenta uma alternativa para o relacionamento de dados dos Sistemas de Informação em Saúde do SUS por meio de técnicas estatísticas e computacionais livres e gratuitas que permitem aos pesquisadores analisar dados de um mesmo usuário em diferentes sistemas, agregando um número maior de variáveis que permitem encontrar padrões de morbidade e mortalidade que podem subsidiar medidas e políticas de promoção, prevenção, diagnóstico, tratamento e reabilitação de agravos e doenças tratadas no sistema público.

Como limitações deste estudo podemos relatar a influência de possíveis erros sistemáticos provenientes da utilização de bancos de dados secundários, como a má qualidade de dados pessoais – que pode ter impactado no total de pares encontrados –; e a perda de alguns resultados verdadeiros-positivos no processo de exclusão automática de falsos-positivos.

## **CONCLUSÕES**

As técnicas de vinculação probabilística de dados permitem o relacionamento de dados que não possuem uma chave de identificação única. Diante de um cenário em que os diferentes SIS não possuem variáveis de identificação para os usuários, o relacionamento probabilístico é uma ferramenta útil para melhor definir o perfil de mortalidade, refinar as análises de situação em saúde e aumentar a precisão de informações relacionadas às causas dos óbitos.

A recuperação de informações epidemiológicas de casos de malária para investigar óbitos pela doença é essencial para subsidiar o PNCM na vigilância de óbitos da doença e no controle de casos de malária grave que podem evoluir a óbito.

É importante que os protocolos de aquisição ou manejo de dados sensíveis sejam melhor descritos no nível federal, estadual e municipal, para

que processos de vinculação de dados sejam possíveis de serem realizados por instituições de ensino e pesquisa além das unidades técnicas dos serviços de saúde. Essa colaboração entre academia e serviços de saúde é essencial para o fortalecimento da saúde pública do país.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. Coelho GC, Chioro A. Afinal, quantos Sistemas de Informação em Saúde de base nacional existem no Brasil?. *Cadernos de Saúde Pública* [online]. 2021, v. 37, n. 7 [Acessado 31 Agosto 2022], e00182119. Disponível em: <<https://doi.org/10.1590/0102-311X00182119>>. Epub 19 Jul 2021. ISSN 1678-4464. <https://doi.org/10.1590/0102-311X00182119>.
2. Enamorado T, Imai K. Validating self-reported turnout by linking public opinion surveys with administrative records. *Public Opinion Quarterly*, v. 83, n. 4, p. 723-748, 2019. Disponível em: <https://imai.fas.harvard.edu/research/files/turnout.pdf>
3. Cerina R, Duch R. Digital Trace and Machine Learning with Post-Stratification to Measure Public Opinion. Disponível em: [https://kops.uni-konstanz.de/bitstream/handle/123456789/36418/Jungherr\\_0-380412.pdf;jsessionid=05F125540AE6AA903400213F9A1B6487?sequence=1](https://kops.uni-konstanz.de/bitstream/handle/123456789/36418/Jungherr_0-380412.pdf;jsessionid=05F125540AE6AA903400213F9A1B6487?sequence=1)
4. Mcveigh BS, Spahn BT, Murray JS. Scaling Bayesian probabilistic *record linkage* with post-hoc blocking: an application to the california great registers. arXiv preprint arXiv:1905.05337, 2019.
5. Kim IS, Stuckatz J, Wolters L. Strategic and Sequential Links between Campaign: Donations and Lobbying. 2020.
6. Garcia KKS, Miranda CB de, Sousa FN e F de. Procedimentos para vinculação de dados da saúde: aplicações na vigilância em saúde. *Epidemiol Serv Saúde* [Internet]. 2022;31(Epidemiol. Serv. Saúde, 2022 31(3). Available from: <https://doi.org/10.1590/S2237-96222022000300004> Avoundjian T, Dombrowski JC, Golden MR, et al.

- Comparing Methods for *Record linkage* for Public Health Action: Matching Algorithm Validation Study. *JMIR Public Health Surveill.* 2020;6(2):e15917. Published 2020 Apr 30. doi:10.2196/15917
7. Sayers A, Ben-Shlomo Y, Blom AW, Steele F. Probabilistic *record linkage*. *Int J Epidemiol.* 2016;45(3):954-964. doi:10.1093/ije/dyv322
  8. Brasil. Ministério da Saúde. Ficha de Notificação Sivep - Malária Atualizado em 30 de jun, 2022. Acessado em: 31 de ago, 2022. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/publicacoes-svs/malaria/ficha-de-notificacao-sivep-malaria-v03-08-2020.pdf/@@download/file/Ficha%20de%20Notifica%C3%A7%C3%A3o%20Sivep-Mal%C3%A1ria%20v03-08-2020.pdf>
  9. Brasil. Ministério da Saúde. Portal Sinan – Malária: Ficha de notificação/investigação individual para malária. Acessado em: 31 de ago, 2022. Disponível em: [http://portalsinan.saude.gov.br/images/documentos/Agravos/Malaria/Malaria\\_v5.pdf](http://portalsinan.saude.gov.br/images/documentos/Agravos/Malaria/Malaria_v5.pdf)
  10. Gill L. Methodology and Applications of Building a National File of Health and Mortality data. *Austrian Journal of Statistics.* Volume 33 (2004), Number 1&2, 101-124.
  11. Malta DC, Duarte EC. Causas de mortes evitáveis por ações efetivas dos serviços de saúde: uma revisão da literatura. *Ciênc saúde coletiva* [Internet]. 2007May;12(Ciênc. saúde coletiva, 2007 12(3):765–76. Available from: <https://doi.org/10.1590/S1413-81232007000300027>
  12. Churches T, Christen P, Lim K, Zhu JX. Preparation of name and address data for *record linkage* using hidden Markov models. *BMC Med Inform Decis Mak.* 2002; 2:9. Doi:10.1186/1472-6947-2-9
  13. Yaohao P, Mation LF. O Desafio Do Pareamento De Grandes Bases De Dados: Mapeamento De Métodos De *Record linkage* Probabilístico E Diagnóstico De Sua Viabilidade Empírica. Instituto de Pesquisa Econômica Aplicada- Brasília: Rio de Janeiro: IPEA. 2018. ISSN 1415-4765.

14. Organização Mundial da Saúde (OMS). Malaria [ONLINE]. Update at July 26, 2022. Accessed: Sep 5, 2022. Disponível em: <https://www.OMS.int/news-room/fact-sheets/detail/malaria#:~:text=Left%20untreated%2C%20P.,a%20period%20of%2024%20hours>
15. Zhang S, Hu Y, Bian G,. Research on string similarity algorithm based on Levenshtein Distance. In: 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). IEEE, 2017. p. 2247-2251.
16. Liao H, Xu Z. Approaches to manage hesitant fuzzy linguistic information based on the cosine distance and similarity measures for HFLTSS and their application in qualitative decision making. *Expert Systems with Applications*, v. 42, n. 12, p. 5328-5336, 2015.
17. David R, Jennifer B, Joran E. fuzzyJoin [Internet]. New York: Datacamp; 2020 [update 2020 May 15; cited 2022 Sep 2]. Available from: <https://cran.r-project.org/web/packages/fuzzyjoin/fuzzyjoin.pdf>
18. Grannis SJ, Overhage JM, Hui S, McDonald CJ. Analysis of a probabilistic *record linkage* technique without human review. *AMIA Annu Symp Proc*. 2003:259-63.
19. Brasil. Presidência da República. Secretaria-Geral. Subchefia para Assuntos Jurídicos. Lei Geral de Proteção de Dados Pessoais (LGPD) - LEI Nº 13.709, DE 14 DE AGOSTO DE 2018. Acessado em: 05 de Set, 2022. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/L13709compilado.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm)
20. Maia LTS, Souza WV, Mendes ACG. A contribuição do linkage entre o SIM e SINASC para a melhoria das informações da mortalidade infantil em cinco cidades brasileiras. *Revista Brasileira de Saúde Materno Infantil* [online]. 2015, v. 15, n. 1 [acessado 2 novembro 2021], pp. 57-66. Disponível em: <<https://doi.org/10.1590/S1519-38292015000100005>>. ISSN 1806-9304. <https://doi.org/10.1590/S1519-38292015000100005>.
21. Rocha MS, Oliveira GP, Guillen LCT, Coeli CM, Saraceni V, Pinheiro RS. Uso de linkage entre diferentes bases de dados para qualificação de

- variáveis do Sinan-TB e a partir de regras de scripting. Cadernos de Saúde Pública [online]. 2019, v. 35, n. 12 [acessado em 3 de novembro 2021], e00074318. Disponível em: <<https://doi.org/10.1590/0102-311X00074318>>. Epub 28 Nov 2019. ISSN 1678-4464. <https://doi.org/10.1590/0102-311X00074318>.
22. Shah GH, Lertwachara K, Ayanso A. *Record linkage* in healthcare: Applications, opportunities, and challenges for public health. International Journal of Healthcare Delivery reform Initiatives (IJHDRI), v. 2, n. 3, p. 29-47, 2010.
23. Brasil. Ministério da Saúde. Departamento de Informática do Sistema Único de Saúde (Datasus). Dicionário de dados - e-SUS Notifica [ONLINE]. Atualizado em 16 de agosto, 2021. Disponível em: [https://datasus.saude.gov.br/wp-content/uploads/2021/08/Dicionario-de-dados\\_E-SUS-Notifica\\_16\\_08\\_21.pdf](https://datasus.saude.gov.br/wp-content/uploads/2021/08/Dicionario-de-dados_E-SUS-Notifica_16_08_21.pdf)
24. Enamorado T, Fifield B, Imai K. FastLink [Internet]. New York: Datacamp; 2020 [update 2020 Apr 29; cited 2022 Sep 2]. Available from: <https://cran.r-project.org/web/packages/fastLink/fastLink.pdf>
25. Camargo Jr., Kenneth R. de e Coeli, Cláudia M. Reclink: aplicativo para o relacionamento de bases de dados, implementando o método probabilistic *record linkage*. Cadernos de Saúde Pública [online]. 2000, v. 16, n. 2 [Acessado 30 Julho 2022], pp. 439-447. Disponível em: <<https://doi.org/10.1590/S0102-311X2000000200014>>. Epub 15 Ago 2001. ISSN 1678-4464. <https://doi.org/10.1590/S0102-311X2000000200014>
26. Nahorniak J, Bovbjerg V, Case S, Kincl L. Application of data linkage techniques to Pacific Northwest commercial fishing injury and fatality data. *Inj Epidemiol.* 2021;8(1):26. Published 2021 Jul 5. doi:10.1186/s40621-021-00323-z

## **5.4. Descrição dos fatores associados ao óbito por malária (Objetivo específico 4)**

### **5.4.1. Fatores associados à morte por malária no Brasil: um estudo de caso-controle de 2011 a 2020**

#### **INTRODUÇÃO**

A malária, uma enfermidade infecciosa aguda caracterizada que prevalece principalmente em nações tropicais e em desenvolvimento. De acordo com a Organização Mundial da Saúde (OMS), a malária afetou cerca de 247 milhões de indivíduos e causou mais de 619.000 óbitos no ano de 2021 [1]. Apesar dos esforços empreendidos em escala mundial para o controle da malária e de suas interligações com os Objetivos do Desenvolvimento Sustentável (ODS) promulgados pela Organização das Nações Unidas (ONU), esta enfermidade, além de sua prevalência notável, exibiu – globalmente - em 2021 uma taxa de mortalidade de 14,8 para 100.000 pessoas em risco. [1].

O agente etiológico da malária é um protozoário pertencente ao gênero *Plasmodium*, transmitido aos seres humanos por meio da picada da fêmea do mosquito vetor. As espécies de *Plasmodium* que afetam os seres humanos incluem *Plasmodium vivax*, *Plasmodium falciparum*, *Plasmodium malariae*, *Plasmodium ovale* e *Plasmodium knowlesi* [2]. No Brasil, as principais espécies envolvidas são *P. vivax* e *P. falciparum*, com ocorrência ocasional de *P. malariae*, que, no entanto, é consideravelmente menos frequente em comparação com *P. vivax* e *P. falciparum*. A principal via de transmissão da malária em áreas endêmicas e não endêmicas é por meio do vetor, cuja ingestão sanguínea contribui para a disseminação da infecção entre os seres humanos [3].

É de suma importância interromper a transmissão da malária a partir de casos já confirmados por meio da administração de medicamentos capazes de eliminar os plasmódios do organismo. O diagnóstico da malária pode ser realizado mediante análise microscópica de esfregaços sanguíneos, por meio da técnica conhecida como gota espessa, que se destaca como o método de diagnóstico mais confiável em comparação com alternativas como testes de diagnóstico rápido e análises moleculares [4]. Uma vez realizado o diagnóstico, torna-se imprescindível instituir um tratamento oportuno e adequado para o paciente, visando interromper a progressão da doença e evitar complicações. No Brasil, atualmente, o Sistema Único de Saúde (SUS) disponibiliza testes de diagnóstico e tratamento para malária de forma gratuita [3,5].

No contexto brasileiro, a malária se configura como uma questão grave para a Saúde Pública, particularmente na região da Amazônia Legal, a qual engloba aproximadamente 99% dos casos, enquanto a região extra-AMZ é responsável por menos de 1% dos novos casos de malária. Em 2021, de acordo com o Programa Nacional de Prevenção e Controle da Malária (PNCM) do Ministério da Saúde do Brasil, foram registrados aproximadamente 139.211 mil casos autóctones e novos de malária no território brasileiro [6].

Em 2019 no Brasil foram registrados 157.452 casos e 37 óbitos, enquanto em 2020 e 2021 foram registrados, respectivamente 51 e 61 óbitos [6]. Para intensificar esses progressos e trabalhar rumo à eliminação da malária, o Programa Nacional de Prevenção e Controle da Malária do Ministério da Saúde publicou o Plano de Eliminação da Malária em 2022. Este plano visa eliminar as mortes relacionadas à malária até 2030 e os casos autóctones da doença no país até 2035 [7].

Não obstante a aparente baixa incidência de óbitos em relação ao número de casos no Brasil, essas mortes representam fragilidades do Sistema Único de Saúde (SUS) brasileiro, uma vez que os óbitos por malária são considerados evitáveis [8]. Na região da Amazônia Legal, a taxa de letalidade

por malária é de 0,02%, enquanto na região extra-AMZ, essa taxa chega a 2,04%, ou seja, 123 vezes maior do que na região AMZ [9].

Apesar da magnitude da malária no território brasileiro, as causas subjacentes à ocorrência de óbitos por malária carecem de evidências científicas robustas. Estudos conduzidos em nível global apontam que os óbitos por malária estão correlacionados com diversos fatores, incluindo sexo, condições médicas preexistentes, comorbidades, coinfeção com o HIV/AIDS, atraso na busca por assistência médica, tratamento não oportuno, tratamento inadequado, histórico de viagens para áreas endêmicas, sazonalidade e infecção por *P. falciparum* [10-12].

Assim, o objetivo desta pesquisa consiste em descrever e analisar os fatores associados aos óbitos por malária no Brasil no período compreendido entre 2011 e 2020.

## **MÉTODOS**

### *Desenho, população, local e período de estudo*

Estudo do tipo caso-controle com utilização de dados secundários. Foram analisados todos os casos e óbitos por malária de pessoas residentes no Brasil. O território de análise é todo o território brasileiro, incluindo região da Amazônia legal que é composta pelos estados do Acre, Amapá, Amazonas, Mato Grosso, Maranhão, Pará, Rondônia, Roraima e Tocantins, e região extra-AMZ, composta pelos demais estados e o Distrito Federal. O período análise será de 2011 a 2020.

### *Fonte e Procedimento para Coleta de Dados:*

Para realização deste estudo fora solicitado acesso a diferentes bancos de dados sob supervisão do Ministério da Saúde do Brasil descrito a seguir. Para construção do banco de dados foi solicitado acesso aos dados provenientes dos seguintes sistemas de informação: Sistema de vigilância epidemiológica da malária (Sivep-Malária) para casos positivos de malária

registrados na região da Amazônia legal; Sistema de Informação de Agravos de Notificação (Sinan) para casos positivos de malária na região extra-AMZ; Sistema de Informação de Mortalidade (SIM) para registros de óbitos por malária. Utilizou-se técnicas de *record linkage* probabilístico e processos específicos de limpeza de dados para aprimorar a eficiência da correspondência de registros, de acordo com as orientações de Garcia et al [13]. Através da ligação, podemos recuperar informações epidemiológicas sobre mortes por malária que foram previamente registradas como casos da doença no Sivep-Malária ou no Sinan. Os dados foram adquiridos em fevereiro de 2022.

#### *Variáveis analisadas*

Neste trabalho foram analisadas as seguintes variáveis epidemiológicas, provenientes dos bancos Sivep-Malária e Sinan: Sexo, Raça/cor, Faixa etária, Tipo de detecção (TIPO\_LAM), Parasitemia em cruzes, Data de notificação, data de início de sintomas, Espécie parasitária, ocupação nos últimos 15 dias, área especial de infecção (disponível apenas no Sivep-malária) e País de infecção. As variáveis provenientes do SIM foram: Local de residência, causa básica do óbito, causas associadas ao óbito e data de óbito.

#### *Vinculação de dados (Probabilistic linkage)*

Para o desenvolvimento do procedimento de vinculação de registros, empregou-se a interface do R Studio, junto com técnicas de vinculação probabilística, conforme detalhado a seguir, utilizando a linguagem de programação R (versão 4.2.1). A vinculação probabilística consiste em uma técnica que se vale de um conjunto de regras combinatórias para avaliar a probabilidade de distintos registros pertencerem ao mesmo indivíduo. Tal avaliação é realizada por meio do cálculo de medidas de similaridade entre variáveis de natureza aberta, categóricas, de data ou, ainda, de medidas de dissimilaridade entre variáveis numéricas [14].

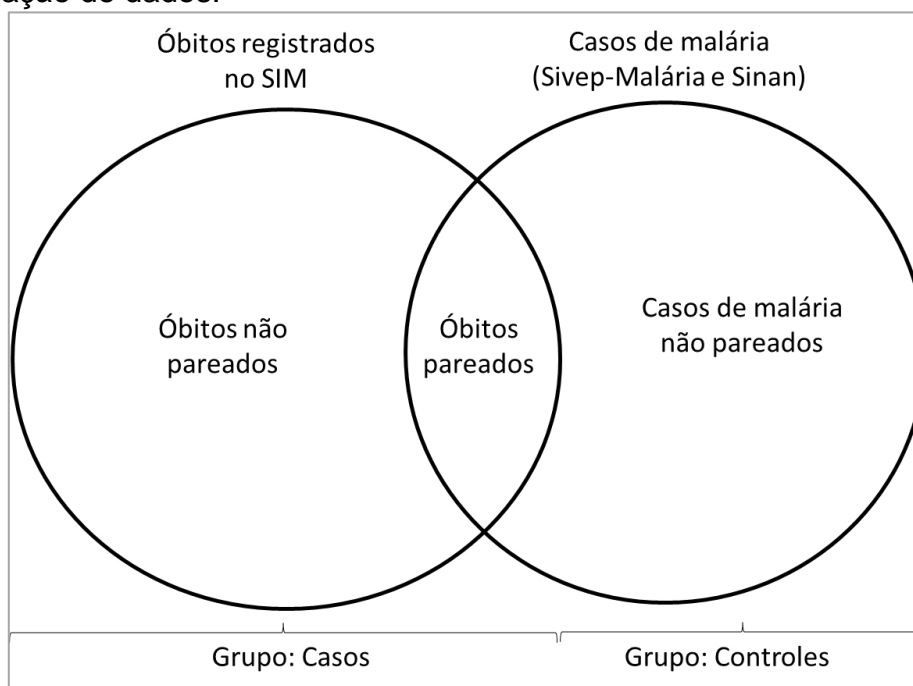
Na etapa de emparelhamento, foram selecionadas as seguintes variáveis-chave de vinculação: "Nome do paciente", "Data de nascimento" e "Nome da mãe", seguindo diretrizes propostas por estudos anteriores [13,15,16]. Para a realização do procedimento de vinculação de registros, empregou-se o pacote "fuzzyjoin" [17].

Nesse contexto, as distâncias de Levenshtein [18] e de Cosseno [19] foram utilizadas. Especificamente, a distância de Levenshtein foi parametrizada com o valor 3, enquanto, para a distância de Cosseno, o valor estabelecido foi de 0,5. Maiores detalhes metodológicos podem ser encontrados no Capítulo 3 da seção de resultados.

#### *Definição de Casos e Controles*

Os casos são definidos como indivíduos registrados no SIM (Sistema de Informações sobre Mortalidade) devido à causa básica ou associada da malária (CID-10: B50 a B54). Os controles consistem em todos os novos casos de malária (excluindo recorrências/notificações de verificação de lâminas de cura) que não possuem pares correspondentes no processo de vinculação (Figura 15).

Figura 15. Definição de casos e controles de acordo com o processo de vinculação de dados.



Fonte: Autoria própria.

#### *Tratamento de dados*

Para realizar a análise de regressão logística, realizou-se um tratamento das variáveis para agregação de variáveis que compartilhavam características similares e a exclusão de variáveis com contagens de valores observados insuficientes para suportar a análise.

A variável "raça/cor negra" foi criada pela combinação das categorias "pardo" e "preto". A variável "brancos" incluiu registros classificados como "brancos" e "amarelos". Além disso, foi introduzida a variável "não-branco", que foi utilizada apenas na regressão logística relacionada a óbitos na região extra-AMZ. Essa variável englobou registros de "pardos", "pretos" e "indígenas". A inclusão dos indígenas nessa categoria foi necessária devido à ausência de óbitos registrados para indígenas na região extra-AMZ.

"Ocupações nos últimos 15 dias" foram categorizadas em 5 categorias diferentes: "agricultura" (agricultura, caça/pesca, exploração vegetal, pecuária), "outros" (outros, motorista, construção de estradas/barragens), "viagem/turismo", "garimpagem/mineração" e "doméstica".

Para espécie parasitária (resultados de exame), considerou-se as seguintes opções de infecções: *Plasmodium falciparum* (2- Pf, 3- Pf+fg; 7- fg); *P. falciparum* + infecção mista (5- Pf+Pv; 6- Pv+fg; 9- Pg+Pm); *Plasmodium vivax* (4- Pv, 11 – Não *falciparum*); e outras espécies (8- Pm; 10- Po). Para a regressão logística a categoria de “Pf + infecções mistas” foi agregada em “*Plasmodium falciparum*”.

A variável de parasitemia em cruces classificada em “2 cruces ou menos” (++ = até 10.000 por mm<sup>3</sup>), “três cruces” (+++ = 10.001-100.000 por mm<sup>3</sup>) e “4 cruces” (++++ = 100.001 por mm<sup>3</sup> ou mais) [5].

Óbitos por malária importada considerou a variável “País de infecção”. Todos os países diferentes de “Brasil” (código: 001) foram classificados como “importados”, e os infectados no “Brasil” como “autóctones”.

### *Análise Estatística*

Realizou-se análises descritivas caracterização, com apresentação da taxas de letalidade (óbitos / total de casos no período\*100). Para identificar fatores de risco e causas associadas, inicialmente realizou-se regressões logísticas univariadas para verificar a Odds-Ratio bruta (Razão de chances bruta) e posteriormente as variáveis foram inseridas no modelo múltiplo seguindo o método de *stepwise* [11,20].

### *Softwares*

Para o manuseio, tratamento e análise dos dados, foram utilizados os seguintes softwares: R versão 4.2.3 [21], Microsoft Excel 2016, Tableau software (versão 2022.3) [22] e Qgis versão 3.28 [23].

### *Considerações Éticas*

Este projeto recebeu aprovação do comitê de ética em pesquisa da Faculdade de Medicina da Universidade de Brasília em janeiro de 2022. A aprovação está documentada sob o protocolo CEP: 5.216.688 (51246121.0.0000.5558).

## RESULTADOS

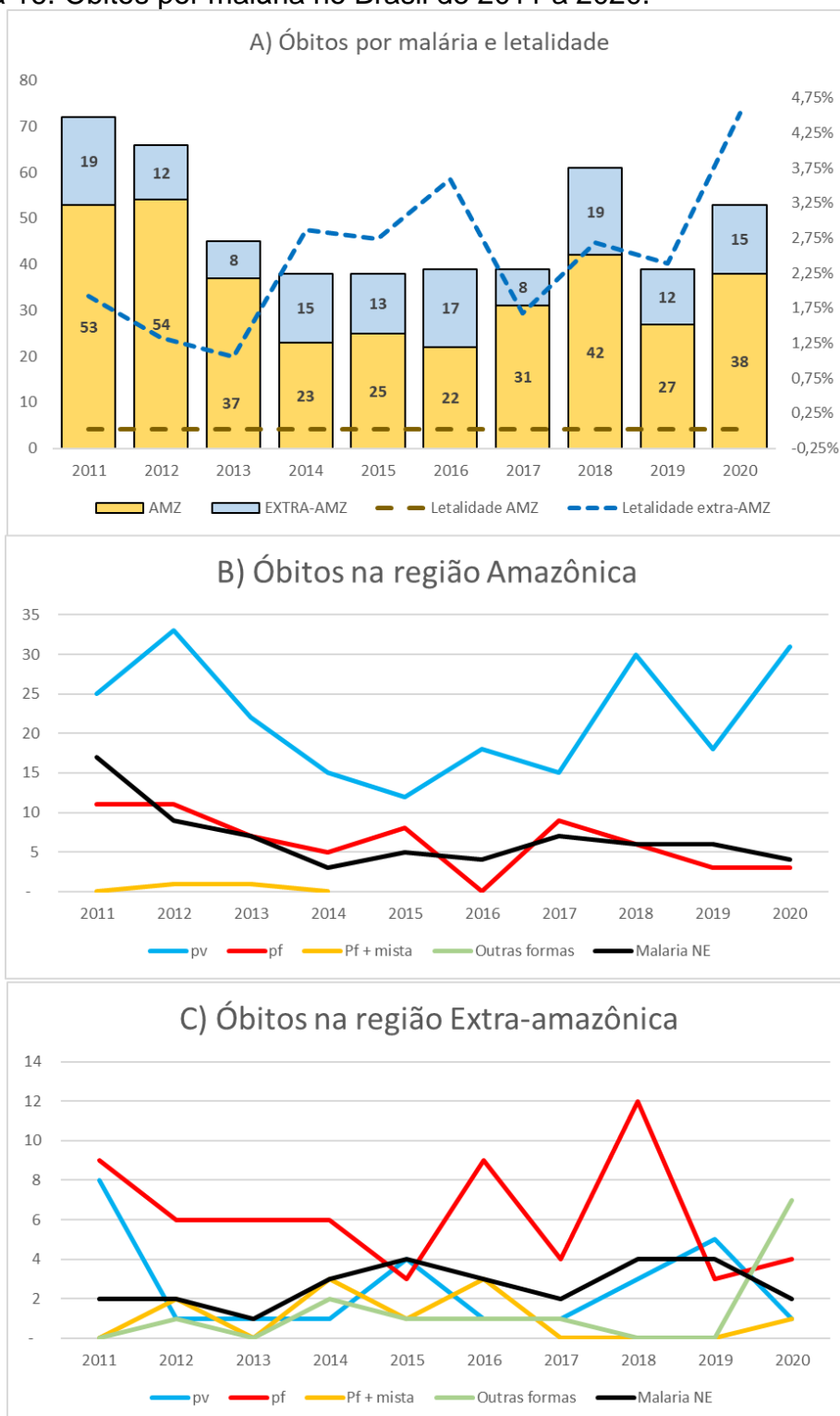
Entre 2011 e 2020 foram registrados 490 óbitos por malária. Destes, 216 (44%) tiveram informações epidemiológicas recuperadas no processo de *record linkage*. O grupo controle foi composto por 1.785.792 casos de novas infecções de malária que não evoluíram a óbitos. Em 2011 foram registrados 72 óbitos e em 2020 foram 53, mostrando uma diminuição de 35,8% de 2020 em relação a 2011. Na região AMZ foram registrados 352 (71,8%) e na região extra-AMZ 138 (28,2%) (Figura 16).

No período estudado, a letalidade de malária na região AMZ foi contínua (0,03%), já na região extra-AMZ a letalidade variou de 1,06% (2013) a 4,53% (2020). A letalidade chegou a ser 207,7% vezes maior na região extra-AMZ em comparação com a região AMZ em 2016, em 2020 esse valor foi de 172,4%. Dentre os 53 óbitos ocorridos em 2020 apenas sete óbitos apresentaram coinfeção com COVID-19 (CID-10: B342; U07).

Dentre os 490 óbitos foram registradas 135 diferentes causas relacionadas além da malária. Dentre as 135, as 3 mais frequentes foram “outras infecções bacterianas não especificadas” (CID-10: A41), “Insuficiência respiratória aguda” (CID-10: J96) e “Sintomas e sinais que envolvem o sistema circulatório” (CID-10: R09). Estas três foram citadas em 229 (46,7%) dos óbitos.

As infecções mais frequentes dentre óbitos foram por *Plasmodium vivax* (N = 245; 50%) seguido por *Plasmodium falciparum* (N= 125; 25,5%), Malária não especificada (CID-10: B54) foi declarada em 95 (19,4%) óbitos (Figura 16).

Figura 16. Óbitos por malária no Brasil de 2011 a 2020.



Legenda: A) Óbitos e letalidade por malária segundo região de residência; B) Óbitos por malária segundo espécie entre residentes da região *Amazônica* C) Óbitos por malária segundo espécie entre residentes da região *extra-Amazônica*. Fonte: Sivep-Malária, Sinan e SIM – Ministério da Saúde.

O perfil epidemiológico destacou: 284 (58,0%) óbitos em indivíduos do sexo masculino; 290 (59,2%) óbitos entre indivíduos negros e 60 (12,2%) óbitos foram entre indígenas; A faixa etária mais afetada foi de 60 ou mais com 149 (30,4%) óbitos, seguido por indivíduos de 20 a 39 anos e de 40 a 59 anos, com respectivamente 122 (24,9%) e 120 óbitos (24,5%). Foram registrados apenas dois óbitos por gestantes (0,4%) (Tabela 6).

As ocupações mais frequentes foram relacionadas a Turismo e Viagens e Agricultura/Rural, cada uma sendo responsável por 7,8% dos óbitos. “Outras” ocupações totalizaram 17,3% dos óbitos e “ignorado/ em branco” totalizou 58,6% do banco. Quanto à área especial de infecção, 65 óbitos (13,3%) foram em área rural e 30 (6,1%) foram em áreas indígenas. Óbitos provenientes de infecções em áreas urbanas contabilizaram apenas 17 óbitos (3,5%). Ao todo 148 (30,2%) óbitos foram decorrentes de casos autóctones do Brasil e 68 (13,9%) foram de casos importados de outro país (Tabela 6).

Um total de 184 (37,6%) óbitos tiveram o diagnóstico de malária feito por detecção passiva e apenas 27 (5,5%) tiveram o diagnóstico por atividades de detecção ativa. Quanto a oportunidade de diagnóstico, 67 (13,7%) óbitos foram diagnosticados com malária em menos de 72 horas após o início de sintomas. Diagnósticos acima de 72h foram registrados em 147 óbitos, representando 30,0% dos óbitos. Quanto à parasitemia, foram 143 (29,2%) óbitos com 2 cruces ou menos ( $\leq ++$ ), 41 (8,4%) com 3 cruces (+++) e 24 (4,9%) com 4 cruces (++++) (Tabela 6).

Observou-se significativas diferenças no perfil epidemiológico entre os óbitos de residentes na região AMZ (AMZ) e aqueles ocorridos na região extra-AMZ. Notavelmente, não houve óbitos de indígenas na região extra-AMZ, porém, houve uma maior proporção de óbitos entre a população branca. A ocorrência de óbitos em indivíduos com menos de 20 anos foi escassa na região extra-AMZ, enquanto na região AMZ, esses casos representaram 26,7% de todos os óbitos registrados (Tabela 6).

A frequência de óbitos cujos casos foram identificados por detecção ativa na região extra-AMZ foi 2,4 vezes maior do que na região AMZ. Além

disso, a prevalência de parasitemia alta (+++) entre os óbitos na região extra-AMZ foi significativamente maior, atingindo 35,2%, enquanto na região AMZ essa taxa foi de apenas 4,3% dos óbitos (Tabela 6).

Outro aspecto notável foi a maior proporção de óbitos com tempo de notificação superior a 72 horas na região extra-AMZ, correspondendo a 49,3% dos casos, enquanto na região AMZ, esse percentual foi de 22,4% (Tabela 6).

Em relação à espécie de *Plasmodium* causadora dos óbitos, *P. falciparum* foi responsável por 52,2% dos óbitos na região extra-AMZ, e ocorreram 13 óbitos (9,4%) devido a outras espécies, como *P. ovale* ou *P. malariae*. Por outro lado, na região AMZ, 62% (n=219) dos óbitos foram causados por *P. vivax* (Tabela 6).

Foi identificada a predominância de óbitos relacionados a casos importados de outros países na região extra-AMZ (n=60, 43,5%), enquanto na região AMZ, esse percentual foi pequeno, representando apenas 2,3% dos óbitos (Tabela 6).

Em relação às ocupações mais frequentes, na região extra-AMZ, elas estavam relacionadas a viagens e turismo (N=34; 24,6%), enquanto na região AMZ, atividades domésticas (N=27; 7,7%) e agricultura (N=27; 7,7%) foram mais comuns. É importante notar que informações sobre a área especial de infecção não estavam disponíveis nos bancos de dados da região extra-AMZ (Tabela 6).

A ocorrência de elevado número de informações ignoradas/em branca se deve ao fato do processo de *record linkage* ter recuperado informações epidemiológicas de apenas 45,0% dos óbitos (Tabela 6).

Tabela 6. Perfil Epidemiológico dos óbitos por malária no Brasil, 2011-2020.

	Região Amazônica (N= 352)		Região extra-Amazônica (N = 138)		Total (N = 490)	
	N	%	N	%	N	%
<b>Sexo</b>	352	71,8%	138	28,2%	490	100,0%
Feminino	149	42,3%	57	41,3%	206	42,0%
Masculino	203	57,7%	81	58,7%	284	58,0%
<b>Raça/cor</b>						
Branca	50	14,2%	66	47,8%	116	23,7%
Negra	229	65,1%	61	44,2%	290	59,2%
Indígena	60	17,0%	0	0,0%	60	12,2%
Ignorado/em branco	13	3,7%	11	8,0%	24	4,9%
<b>Faixa etária</b>						
Menor de 6 meses	19	5,4%	0	0,0%	19	3,9%
De 06 a 11 meses	8	2,3%	0	0,0%	8	1,6%
01 a 04 anos	21	6,0%	2	1,4%	23	4,7%
05 a 09 anos	15	4,3%	0	0,0%	15	3,1%
10 a 19 anos	31	8,8%	2	1,4%	33	6,7%
20 a 39 anos	80	22,7%	42	30,4%	122	24,9%
40 a 59 anos	77	21,9%	43	31,2%	120	24,5%
60 ou mais	100	28,4%	49	35,5%	149	30,4%
Ignorado/em branco	1	0,3%	0	0,0%	1	0,2%
<b>Tipo de detecção</b>						
Detecção Passiva	122	34,7%	62	44,9%	184	37,6%
Detecção Ativa	14	4,0%	13	9,4%	27	5,5%
Ignorado/em branco	216	61,4%	63	45,7%	279	56,9%
<b>Parasitemia em cruzes (mm<sup>3</sup>)</b>						
< ++ (< 9,999 mm <sup>3</sup> )	113	32,1%	30	21,7%	143	29,2%
+++ (10,000 - 100,000 mm <sup>3</sup> )	9	2,6%	32	23,2%	41	8,4%
++++ (>100,000 mm <sup>3</sup> )	6	1,7%	18	13,0%	24	4,9%
Ignorado/em branco	224	63,6%	58	42,0%	282	57,6%
<b>Oportunidade de detecção</b>						
Menor que 72h	55	15,6%	12	8,7%	67	13,7%
Maior que 72h	79	22,4%	68	49,3%	147	30,0%
Ignorado/em branco	218	61,9%	58	42,0%	276	56,3%
<b>Espécie parasitária</b>						
<i>P. vivax</i>	219	62,2%	26	18,8%	245	50,0%
<i>P. falciparum</i> + outra espécie	65	18,5%	72	52,2%	137	28,0%
Outras formas	0	0,0%	13	9,4%	13	2,7%
Malária Não especificada	68	19,3%	27	19,6%	95	19,4%
<b>Ocupação nos últimos 15 dias</b>						
Agricultura/Rural	32	9,1%	6	4,3%	38	7,8%
Doméstica	27	7,7%	4	2,9%	31	6,3%
Garimpagem ou Mineração	8	2,3%	3	2,2%	11	2,2%
Outros	59	16,8%	26	18,8%	85	17,3%
Turismo e Viagens	4	1,1%	34	24,6%	38	7,8%
Ignorado/em branco	222	63,1%	65	47,1%	287	58,6%

Ignorado/em branco	222	63,1%	65	47,1%	287	58,6%
<b>Área de infecção</b>						
Aldeia indígena	30	8,5%	0	0,0%	30	6,1%
Urbana	17	4,8%	0	0,0%	17	3,5%
Rural	65	18,5%	0	0,0%	65	13,3%
Garimpo	9	2,6%	0	0,0%	9	1,8%
Ignorado/em branco	231	65,6%	138	100,0%	369	75,3%
<b>País de infecção</b>						
Autóctone do Brasil	128	36,4%	20	14,5%	148	30,2%
Importado de outro país	8	2,3%	60	43,5%	68	13,9%
Ignorado/em branco	216	61,4%	58	42,0%	274	55,9%

Legenda: Fonte: Sivep-Malária, Sinan e SIM – Ministério da Saúde.

Na análise de regressão logística múltipla de residentes da região AMZ, a variável de detecção ativa teve um nível de proteção de 57% (AOR=0.42; IC 95%: 0.22 - 0.83; p-valor = 0.01197). Como fatores de risco para óbito, foram identificadas as seguintes categorias: "Idade menor de 6 meses" (AOR= 13.09; IC 95%: 4.13 - 41.49; p-valor < 0.0001), "de 06 a 11 meses" (AOR= 8.23; IC 95%: 2.27 - 29.81; p-valor = 0.00134), "de 40 a 59 anos" (AOR= 2.83; IC 95%: 1.37 - 5.86; p-valor = 0.00505) e "60 ou mais" (AOR= 7.96; IC 95%: 3.68 - 17.23; p-valor < 0.0001); ter sido infectado em área indígena (AOR= 2.92; IC 95%: 1.29 - 6.66; p-valor = 0.01053), ter sido notificado após 72 horas após o início dos sintomas (AOR= 1.76; IC 95%: 1.18 - 2.62; p-valor = 0.00533) e ter parasitemia com três cruces (+++) (AOR= 2.82; IC 95%: 1.36 - 5.86; p-valor = 0.00547) ou quatro cruces (++++) (AOR= 28.13; IC 95%: 8.73 - 90.61; p-valor < 0.0001) (Tabela 7).

Na análise de regressão logística múltipla de residentes da região extra-AMZ, identificaram-se os seguintes fatores de risco: ter infecção por *P. falciparum* (AOR= 2.87; IC 95%: 1.22 - 6.77; p-valor = 0.01592), ter infecção por "outras espécies" (*P. ovale* ou *P. malariae*) (AOR= 6.18; IC 95%: 1.18 - 32.41; p-valor = 0.03114), ter parasitemia de três cruces (+++) (AOR= 5.54; IC 95%: 2.91 - 10.56; p-valor < 0.0001) ou quatro cruces (++++) (AOR= 6.41; IC 95%: 2.97 - 13.85; p-valor < 0.0001) e ter contraído malária em outro país (AOR= 2.57; IC 95%: 1.10 - 6.00; p-valor = 0.02906) (Tabela 8).

1 Tabela 7. Regressão logística múltipla de óbitos por malária na região Amazônica brasileira, 2011-2020.

	Casos (N = 352)	Controles (N = 1.779.795)	COR (univariada)	Intervalos de confiança 95%		P-valor	AOR (múltipla)	Intervalos de confiança 95%		P-valor
<b>Sexo</b>										
Masculino [REF]	203	1.078.624								
Feminino	149	701.138	1,13	0,91	1,4	0,26	1,36	0,90	2,05	0,14
<b>Raça/cor</b>										
Branca [REF]	50	126.630								
Negra	229	1.152.159	0,50	0,37	0,68	< 0,01*	1,15	0,56	2,39	0,70
Indígena	60	310.922	0,49	0,34	0,71	< 0,01*	0,53	0,19	1,50	0,23
<b>Faixa etária</b>										
2-Menor de 6 meses	19	11.152	22,93	12,95	40,60	< 0,01*	13,09	4,13	41,49	< 0,01*
3-De 06 a 11 meses	8	13.617	7,91	3,63	17,21	< 0,01*	8,23	2,27	29,81	< 0,01*
4-01 a 04 anos	21	164.947	1,71	0,98	2,98	0,06	2,16	0,89	5,27	0,08
5-05 a 09 anos	15	206.435	0,98	0,53	1,81	0,94	1,00	0,35	2,88	0,99
1-10 a 19 anos [REF]	31	417.248								
6-20 a 39 anos	80	608.269	1,77	1,17	2,68	< 0,01*	1,78	0,89	3,57	0,10
7-40 a 59 anos	77	285.173	3,63	2,40	5,51	< 0,01*	2,83	1,37	5,86	< 0,01*
8-60 ou mais	100	72.806	18,49	12,36	27,66	< 0,01*	7,96	3,68	17,23	< 0,01*
<b>Ocupação nos últimos 15 dias</b>										
Outros [REF]	59	903.651								
Agricultura/Rural	32	445.861	1,10	0,71	1,69	0,66	0,84	0,50	1,41	0,49
Doméstica	27	180.284	2,29	1,45	3,62	< 0,01*	1,61	0,94	2,77	0,08
Garimpagem ou Mineração	8	104.009	1,18	0,56	2,47	0,66	0,68	0,24	1,95	0,47
Turismo e Viagens	4	36.645	1,67	0,61	4,60	0,31	0,40	0,05	2,89	0,36
<b>Área de infecção</b>										
Urbana [REF]	17	258.807								
Garimpo	9	98.178	1,40	0,62	3,13	0,41	1,23	0,47	3,17	0,67
Aldeia indígena	30	307.869	1,48	0,82	2,69	0,19	2,92	1,29	6,66	0,01*
Rural	65	1.057.184	0,94	0,55	1,60	0,80	0,83	0,48	1,44	0,50
<b>Tipo de detecção</b>										
Deteção Passiva [REF]	122	1.327.671								
Deteção Ativa	14	452.124	0,34	0,19	0,59	< 0,01*	0,42	0,22	0,83	0,01*
<b>Oportunidade de detecção</b>										
Menor que 72h [REF]	55	1.021.404								
Maior que 72h	79	649.635	2,26	1,6	3,19	< 0,01*	1,76	1,18	2,62	< 0,01*
<b>Espécie parasitária</b>										
<i>Plasmodium vivax</i> [REF]	219	1.542.777								
<i>Plasmodium falciparum</i> + outra espécie	65	236.616	1,94	1,47	2,55	< 0,01*	1,12	0,64	1,97	0,68
<b>Parasitemia em cruzes (mm<sup>3</sup>)</b>										
< ++ (< 9,999 mm <sup>3</sup> ) [REF]	113	1.668.419								
+++ (10,000 - 100,000 mm <sup>3</sup> )	9	45.671	2,91	1,48	5,74	< 0,01*	2,82	1,36	5,86	< 0,01*
++++ (>100,000 mm <sup>3</sup> )	6	1.756	50,45	22,16	114,83	< 0,01*	28,13	8,73	90,61	< 0,01*

2 Legenda: Pseudo-R<sup>2</sup>: 0.04904048. AIC: 2078.667. BIC: 2357.024. COR: Odds-Ratio Bruta; AOR: Odds-Ratio Ajustada.

3 Fonte: Sivep-Malária, Sinan e SIM – Ministério da Saúde. \*P-valor < 0,05

4

5 Tabela 8. Regressão logística múltipla de óbitos por malária na região extra-Amazônica brasileira, 2011-2020.

	Casos (N = 138)	Controles (N = 5,997)	COR (univariada)	Intervalos de confiança 95%		P-valor	AOR (múltipla)	Intervalos de confiança 95%		P-valor
<b>Sexo</b>										
Masculino [REF]	81	4.762								
Feminino	57	1.235	2,71	1,92	3,83	< 0,01*	1,48	0,82	2,68	0,19
<b>Raça/cor</b>										
Não-branca [REF]	61	2.899								
Branca	66	2.551	1,23	0,86	1,75	0,25	1,28	0,75	2,20	0,36
<b>Faixa etária</b>										
00 a 10 anos	2	168	1,99	0,28	14,28	0,49	1,46	0,09	24,86	0,79
10 a 19 anos [REF]	2	335								
20 a 39 anos	42	2.868	2,45	0,59	10,17	0,21	2,25	0,30	17,19	0,43
40 a 59 anos	43	2.115	3,41	0,82	14,12	0,09	2,29	0,30	17,68	0,42
60 ou mais	49	511	16,06	3,88	66,46	< 0,01*	5,66	0,70	45,83	0,10
<b>Ocupação nos últimos 15 dias</b>										
Outros	26	2.200								
Agricultura/Rural	6	634	0,80	0,33	1,95	0,62	0,99	0,33	2,99	0,98
Doméstica	4	180	1,88	0,65	5,45	0,24	2,73	0,71	10,46	0,14
Garimpagem ou Mineração	3	681	0,37	0,11	1,24	0,10	0,21	0,03	1,59	0,13
Turismo e Viagens	34	1.690	1,70	1,02	2,85	0,04*	1,27	0,71	2,26	0,41
<b>Tipo de detecção</b>										
Detecção Passiva [REF]	62	4.459								
Detecção Ativa	13	1.538	0,61	0,33	1,11	0,10	0,65	0,31	1,35	0,24
<b>Oportunidade de detecção</b>										
Menor que 72h [REF]	12	1.330								
Maior que 72h	68	4.600	1,64	0,88	3,04	0,11	1,65	0,76	3,56	0,20
<b>Espécie parasitária</b>										
<i>Plasmodium vivax</i> [REF]	26	3.978								
Outras formas	13	85	23,40	11,63	47,10	< 0,01*	6,18	1,18	32,41	0,03*
<i>Plasmodium falciparum</i> + outra espécie	72	1.934	5,70	3,63	8,95	< 0,01*	2,87	1,22	6,77	0,01*
<b>Parasitemia em cruzes (mm<sup>3</sup>)</b>										
< ++ (< 9,999 mm <sup>3</sup> ) [REF]	30	5.051								
+++ (10,000 - 100,000 mm <sup>3</sup> )	32	694	7,76	4,69	12,86	< 0,01*	5,54	2,91	10,56	< 0,01*
++++ (>100,000 mm <sup>3</sup> )	18	252	12,03	6,61	21,87	< 0,01*	6,41	2,97	13,85	< 0,01*
<b>País de infecção</b>										
Autóctone do Brasil [REF]	20	3.938								
Importado de outro país	60	2.059	5,74	3,45	9,54	< 0,01*	2,57	1,10	6,00	0,02*

6 Legenda: Pseudo-R<sup>2</sup>: R<sup>2</sup>: 0.2243347. AIC: 553.5943. BIC: 670.7447. COR: Odds-Ratio Bruta; AOR: Odds-Ratio Ajustada.

7 Fonte: Sivep-Malária, Sinan e SIM – Ministério da Saúde. \*P-valor < 0,05

## DISCUSSÃO

Este estudo destacou as principais características dos óbitos por malária entre residentes brasileiros, explicitando a alta ocorrência de óbitos por *P. vivax* na região endêmica (AMZ) e maior ocorrência de óbitos por *P. falciparum* na região não endêmica (extra-AMZ), com um aumento da letalidade por malária na região não endêmica nos últimos anos. Em números absolutos, a ocorrência de óbitos de residentes da região AMZ é maior que entre óbitos de residentes da extra-AMZ, e a maior concentração destes óbitos se dá principalmente nos estados do Pará, Roraima e Rondônia.

Os óbitos foram predominantes em pessoas do sexo masculino e entre pessoas acima de 60 anos de idade. Os casos que evoluíram a óbito foram em sua maioria detectados passivamente e com mais de 72h de início dos sintomas, com parasitemia de “duas cruces” (++) ou menos. Os principais fatores de risco encontrados para residentes da região amazônica foi ter menos de 1 ano de idade, mais de 40 anos, ser infectado em aldeias indígenas, ter detecção da infecção tardiamente (>72h) e ter parasitemia alta. Ter sua infecção detectada de forma ativa mostrou-se um fator protetivo para o óbito. Já na região extra-AMZ foram identificados como fatores de risco para o óbito infecções por *P. falciparum* ou outras espécies diferentes de *P. vivax* (*P. ovale* ou *P. malariae*), ter alta parasitemia (acima de duas cruces) ou ter sido infectado por malária em outro país fora do Brasil.

Os fatores de risco encontrados neste trabalho estão de acordo com o que já se conhece de relatos históricos sobre fatores associados a malária grave como: Atraso no Diagnóstico e Tratamento [24], tipo de *Plasmodium* (*P. falciparum*, *P. ovale* e *P. malariae*) [25, 26], idade e sistema imune mais frágil [27], infecção provenientes de países ou locais endêmicos [28].

A identificação de infecção em aldeias indígenas como fator de risco reflete, no Brasil, os desafios únicos para a gestão e o controle da doença nestas áreas. Populações indígenas se tornam vulneráveis devido a seu isolamento geográfico, dificuldades de acesso aos serviços de saúde, barreiras culturais, saneamento básico, entre outros [29]. A presença da

Secretaria de Saúde Indígena (Sesai) [30] no território encontra obstáculos que podem estar a atrasar o diagnóstico e conseqüentemente o tratamento [32].

Além disso, a proximidade dessas áreas à floresta amazônica e a atividades como o desmatamento e garimpo de ouro contribuem para um ambiente propício à proliferação dos mosquitos vetores da malária [2]. Isso, somado ao diagnóstico tardio, pode aumentar o risco de óbito por malária nas populações indígenas, exigindo intervenções multissetoriais que abordem questões de saúde, ambiente e socioeconomia.

Para reduzir o risco de óbito por malária em áreas indígenas, são necessárias intervenções que abordem não apenas questões de saúde, mas também fatores ambientais, socioeconômicos e culturais [33]. Isso pode incluir o fortalecimento da infraestrutura de saúde nas áreas indígenas, a melhoria do acesso aos serviços de saúde, a educação sobre prevenção e tratamento da malária e o monitoramento do desmatamento e das atividades de garimpo [34].

Definir estratégias que tenham como objetivo proteger as populações com as características identificadas neste estudo podem ajudar a diminuir o número de óbitos e eliminar óbitos pela doença até 2030. Assim, a vigilância e investigação do óbito devem ser fortalecidas, estratégias de autópsia verbal podem ser incorporadas pelos serviços de vigilância epidemiológica para ter maior clareza das causas relacionadas ao óbito [10,5].

O *record linkage* entre os SIS pode desempenhar um papel crucial na vigilância do óbito por malária, assim como já foi utilizado em outros estudos [36, 37], mesmo quando os sistemas de informação em saúde não possuem interoperabilidade [38]. Ao estabelecer conexões entre os registros de casos de malária e registros de óbitos, podem ser identificadas correlações significativas que contribuem para uma melhor compreensão e monitoramento dos eventos de óbito por malária, principalmente em territórios menores como estados e municípios.

O uso de *record linkage* permite também a identificação de padrões geográficos, demográficos e temporais relacionados ao óbito por malária com

informação qualificada, principalmente frente às diversas causas de morte mal definidas (a exemplo, CID-10: B54). Para a implantação de rotinas de vinculação de dados, antes é necessário desenvolver protocolos de compartilhamento de dados, estabelecer padrões de qualidade de dados e garantir a segurança e a privacidade das informações dos pacientes.

Ainda, a necessidade de uma maior qualificação dos serviços de registro de óbitos é fundamental para evitar causas de malária mal especificadas (CID-10 B54) e garantir que os óbitos por malária sejam registrados com maior precisão. A malária é uma doença que, em muitos casos, pode ser assintomática ou apresentar sintomas inespecíficos, o que pode tornar o registro adequado da causa da morte desafiador [39].

É importante investir em: treinamento adequado (os profissionais de saúde responsáveis pelo registro de óbitos devem receber treinamento adequado para reconhecer e registrar corretamente a malária como causa de morte, em áreas endêmicas e não-endêmicas); melhoria na documentação clínica (é importante melhorar a documentação clínica nos casos de malária, detalhando sintomas, exames e resultados laboratoriais para que a causa da morte seja mais precisamente atribuído); educação e sensibilização (a conscientização tanto entre os profissionais de saúde quanto entre a população em geral sobre a gravidade da malária e a importância de relatar casos e óbitos é fundamental para melhorar a qualidade dos registros) e Integração de dados.

A malária causada por *P. vivax* é geralmente considerada menos grave do que a malária causada pelo *P. falciparum* [40, 41] O *P. vivax* é frequentemente associado a formas mais brandas da doença e menos óbitos em comparação ao *P. falciparum*. No entanto, isso não significa que não ocorram óbitos por *P. vivax*. No Brasil, onde há um sistema de vigilância epidemiológica bem estruturado e disponibilidade de diagnóstico e tratamento gratuitos pelo SUS (National Health System) [5], óbitos por *P. vivax* não deveriam ser tão frequentes quanto são.

Óbitos por *P. vivax* podem estar ocorrendo por falta de vigilância epidemiológica qualificada para a identificação e tratamento oportunos desse

caso, mas também por desassistência ou falta de acesso aos serviços de saúde. Ainda, o cenário pode ser pior, pois dentre os 216 casos de malária que foram vinculados no *record linkage* houve 32 óbitos (48,5%) classificados como malária não especificada (CID-10: B54) na declaração de óbito e que no Sivep-malária foram notificados como casos de infecção por *P. vivax*. Então, considerando que 19,4% (N = 95) dos óbitos foram por Malária não especificada (CID-10: B54) é possível que cerca de metade deles seja por *P. vivax*, o que aumentaria a estimativa de óbitos por *P. vivax* do estudo em cerca de aproximadamente 17%.

É possível que o fortalecimento da vigilância, com o aumento da detecção ativa e oportuna de casos de malária possa prevenir óbitos. Dessa forma, amplificar e fortalecer a vigilância epidemiológica, além de controlar e reduzir a quantidade de casos, irá diminuir o número de óbitos pela doença.

Ações de identificação de casos de malária precisam ser urgentemente fortalecidas na região extra-AMZ do Brasil, pois os casos nesta região não são rapidamente identificados por não ser área endêmica da doença. Porém, já alertara o risco de atividades de viagens a áreas com transmissão ativa poderem desenvolver formas grave da doença [12, 42]. E na região extra-AMZ o perfil de casos [13] e óbitos é de pessoas que visitam áreas endêmicas, infectadas por *P. falciparum*, e estas atividades estão relacionadas à suas ocupações. Logo, é preciso também ter um olhar para as pessoas acometidas por malária de uma perspectiva da saúde do trabalhador, para fortalecimento de uma vigilância transdisciplinar.

Esperava-se que atividades de garimpo fossem ser apontadas como fatores de risco para o desenvolvimento de óbito por malária, porém esta informação é proveniente do banco de dados do Sivep-malária e Sinan, e os dados que foram vinculados no processo de *record linkage* identificou apenas 8 óbitos por garimpeiros. Devido às limitações do processo de vinculação de dados, também à sensibilidade que é se auto identificar como um garimpeiro no processo da notificação de infecção por malária, é possível que o banco gerado não tenha sido capaz de identificar todos os óbitos entre garimpeiros,

e por isso esta população não foi indicada como de risco aumentado para o óbito.

Considerando o exposto e que a literatura já identificou que áreas de garimpo aumentam a transmissão de malária [32], sugere-se também investigações desenhadas especificamente para esta população, para entender se trabalhar em áreas de garimpo (como garimpeiros ou outra profissão) pode ter risco aumentado de óbito.

A inserção de variáveis de "Evolução do caso" e "Presença de comorbidade/complicação" nos registros do Sistema de Informação de Vigilância Epidemiológica da Malária (Sivep-Malária) é a principal indicação deste estudo. Pois a realização dele muito se respalda pela ausência de informações de seguimento temporal do caso de malária.

Essas variáveis permitirão análises mais abrangente da progressão das infecções por malária e a identificação de complicações associadas a condições de saúde pré-existentes ou simultâneas. Essa melhoria na documentação e no acompanhamento de casos de malária é fundamental para aprimorar a vigilância epidemiológica e fortalecer a capacidade de resposta às necessidades de saúde pública relacionadas a essa doença.

### *Limitações*

A performance do processo de *record linkage* depende da qualidade da informação dos bancos de dados utilizados [13]. Assim, apesar do método de *record linkage* probabilístico encontrar pares onde os nomes ou datas de nascimentos não são idênticas, o método sofre influência direta da qualidade do sistema. Assim, é importante que ações de qualificação dos serviços de vigilância sejam fortalecidas no sentido de se registrar os nomes do paciente, da mãe e a data de nascimento com maior rigor.

Também, devido ao processo de *record linkage* ter recuperado informações de apenas 45% dos óbitos, é possível que dentro do grupo controle haja casos de malária que tenham evoluído a óbito. Porém, esta parcela corresponde a menos de 0,02% da magnitude de controles, logo,

assume-se que a possível interferência é ínfima. Além disso, é possível que dentre estes óbitos que não encontraram pares no processo de *record linkage* seja devido a inexistência de notificações de casos de malária, ou seja, subnotificação de casos de malária que evoluíram a óbito no Sivep-Malária.

Neste estudo não foi possível analisar óbitos associados a recorrências de malária [43] devido à falta de informação disponível sobre este evento nos bancos de dados utilizados. O mesmo se aplica a ausência de informações sobre deficiências de G6PD [44] e sobre comorbidades (no grupo controle) [45-47]. A presença de comorbidades, como doenças crônicas, pode influenciar significativamente a gravidade da malária e o risco de óbito. A ausência desses dados no grupo controle dificulta a análise comparativa entre os grupos de casos e controles em relação às comorbidades e seu impacto nos desfechos de óbito por malária.

## CONCLUSÕES

A vigilância do óbito e sua investigação são ações que precisam urgentemente serem fortalecidas no território nacional brasileiro. A implementação de rotinas de vinculação de dados contribuirá para este processo enquanto um sistema de informação em saúde integrado não for desenvolvido no âmbito do Ministério da Saúde. A inserção de variáveis de acompanhamento como “evolução do caso” e “comorbidades” na ficha de investigação epidemiológica também fortalecerá esta agenda.

fatores ambientais, sociais, econômicos, clínicos epidemiológicos podem estar relacionados aos óbitos por espécies de *P. vivax*. Apesar da baixa letalidade de malária no território nacional, estes são óbitos evitáveis que podem ser controlados com o fortalecimento de ações de vigilância epidemiológica.

O número de óbitos por malária no Brasil tem aumentado, mesmo com as subnotificações destes registros. Ainda assim, é possível sim que a meta de se eliminar óbitos por malária no Brasil até 2030 seja alcançado. Porém,

são necessárias ações multisetoriais com forte investimento econômico e interesse político para isso.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. World Health Organization. World malaria report 2022. Genebra. 2022.
2. Talapko J, et al. Malária: O passado e o presente. Microorganisms. MDPI AG, jun. 2019.
3. Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Articulação Estratégica de Vigilância em Saúde. Guia de Vigilância em Saúde. 5th ed. Brasília: MS; 2022.
4. Shahbodaghi SD, Rathjen NA. Malaria: Prevention, Diagnosis, and Treatment. Am Fam Physician. 2022 Sep;106(3):270-278. PMID: 36126008.
5. Brasil. Guia de tratamento da malária no Brasil. Ministério da Saúde do Brasil, Secretaria de Vigilância em Saúde, Departamento de Imunização e Doenças Transmissíveis. Brasília: Ministério da Saúde do Brasil, 2020.
6. Brasil. Panorama epidemiológico da malária em 2021: buscando o caminho para a eliminação da malária no Brasil – Boletim epidemiológico. Coordenação-Geral de Vigilância de Zoonoses e Doenças de Transmissão Vetorial do Departamento de Imunização e Doenças Transmissíveis da Secretaria de Vigilância em Saúde. 2022. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/m/malaria/situacao-epidemiologica-da-malaria/boletins-epidemiologicos-de-malaria/boletim-epidemiologico-vol-53-no17-2022-panorama-epidemiologico-da-malaria-em-2021-buscando-o-caminho-para-a-eliminacao-da-malaria-no-brasil/view>. Acessado em 03 de outubro de 2023.
7. Brazil. Elimina Malária Brasil: Plano Nacional de Eliminação da Malária (National Malaria Elimination Plan). Secretaria de Vigilância em Saúde. Ministério da Saúde. 2022. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/svsa/malaria/politicas-de->

saude/elimina-malaria-brasil-plano-nacional-de-eliminacao-da-malaria/@ @download/file

8. Malta DC, Duarte EC. Causas de mortes evitáveis por ações efetivas dos serviços de saúde: uma revisão da literatura. *Ciênc saúde coletiva* [Internet]. 2007 May;12(Ciênc. saúde coletiva, 2007 12(3):765–76. Available from: <https://doi.org/10.1590/S1413-81232007000300027>
9. Brasil. Ministério da Saúde do Brasil. Secretaria de Vigilância em Saúde, Departamento de Imunizações e Doenças Transmissíveis, Coordenação Geral de Vigilância de Zoonoses e Doenças Transmissíveis por Vetores. *Boletim Epidemiológico – Malária: 2020* [Internet]. novembro de 2020. p. 118. Disponível em: [https://www.gov.br/saude/ptbr/assuntos/media/pdf/2020/dezembro/03/boletim\\_especial\\_malaria\\_1dez20\\_final.pdf](https://www.gov.br/saude/ptbr/assuntos/media/pdf/2020/dezembro/03/boletim_especial_malaria_1dez20_final.pdf)
10. Deressa W, Fantahun M, Ali A. Malaria-related mortality based on verbal autopsy in an area of low endemicity in a predominantly rural population in Ethiopia. *Malar J.* 2007;6:128. Published 2007 Sep 21. doi:10.1186/1475-2875-6-128
11. Sarkar J, et al. Fatores de risco para óbitos por malária no distrito de Jalpaiguri, Bengala Ocidental, Índia: Evidência para ação adicional. *Malaria Journal*, v. 8, n. 1, 2009.
12. Lüthi B, Schlegelhauf P. Fatores de risco associados às mortes por malária em viajantes: uma revisão da literatura. *Travel Medicine and Infectious Disease.* Elsevier USA, jan. 2015.
13. Garcia KKS, Miranda CB de, Sousa FN e F de. Procedimentos para vinculação de dados da saúde: aplicações na vigilância em saúde. *Epidemiol Serv Saúde* [Internet]. 2022;31(Epidemiol. Serv. Saúde, 2022 31(3)). Available from: <https://doi.org/10.1590/S2237-96222022000300004>
14. Sayers A, Ben-Shlomo Y, Blom AW, Steele F. Probabilistic *record linkage*. *Int J Epidemiol.* 2016;45(3):954-964. doi:10.1093/ije/dyv322
15. Churches T, Christen P, Lim K, Zhu JX. Preparation of name and address data for *record linkage* using hidden Markov models. *BMC Med Inform Decis Mak.* 2002; 2:9. Doi:10.1186/1472-6947-2-9

16. Yaohao P, Mation LF. O Desafio Do Pareamento De Grandes Bases De Dados: Mapeamento De Métodos De *Record linkage* Probabilístico E Diagnóstico De Sua Viabilidade Empírica. Instituto de Pesquisa Econômica Aplicada- Brasília: Rio de Janeiro: IPEA. 2018. ISSN 1415-4765
17. David R, Jennifer B, Joran E. fuzzyJoin [Internet]. New York: Datacamp; 2020 [update 2020 May 15; cited 2022 Sep 2]. Available from: <https://cran.r-project.org/web/packages/fuzzyjoin/fuzzyjoin.pdf>
18. Liao H, Xu Z. Approaches to manage hesitant fuzzy linguistic information based on the cosine distance and similarity measures for HFLTSSs and their application in qualitative decision making. *Expert Systems with Applications*, v. 42, n. 12, p. 5328-5336, 2015.
19. David R, Jennifer B, Joran E. fuzzyJoin [Internet]. New York: Datacamp; 2020 [update 2020 May 15; cited 2022 Sep 2]. Available from: <https://cran.r-project.org/web/packages/fuzzyjoin/fuzzyjoin.pdf>
20. Meireles BM, et al. Fatores associados à malária em populações indígenas: um estudo retrospectivo de 2007 a 2016. *PLOS ONE*, v. 15, n. 10, p. e0240741, 21 out. 2020.
21. R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
22. Tableau Software. Tableau Desktop. Versão 2021.3. [Software]. 2021. Disponível em: <https://www.tableau.com/>. Acessado em 8 de outubro de 2023.
23. QGIS Development Team. QGIS Geographic Information System. Versão 3.16. [Software]. 2020. Disponível em: <https://www.qgis.org/>. Acessado em 8 de outubro de 2023.
24. Bastaki H, Carter J, Marston L, Cassell J, Rait G. Time delays in the diagnosis and treatment of malaria in non-endemic countries: A systematic review. *Travel Med Infect Dis*. 2018 Jan-Feb;21:21-27. doi: 10.1016/j.tmaid.2017.12.002. Epub 2017 Dec 7. PMID: 29225124.
25. Hwang J, Cullen KA, Kachur SP, Arguin PM, Baird JK. Severe morbidity and mortality risk from malaria in the United States, 1985-2011. *Open*

- Forum Infect Dis. 2014;1(1):ofu034. Published 2014 Jun 30. doi:10.1093/ofid/ofu034.
26. Legros F, Bouchaud O, Ancelle T, Arnaud A, Cojean S, Le Bras J, et al. Risk factors for imported fatal Plasmodium falciparum malaria, France, 1996-2003. *Emerg Infect Dis* 2007;13:883-8.
  27. Mühlberger N, Jelinek T, Behrens RH, Gjørup I, Coulaud JP, Clerinx J, et al. Age as a risk factor for severe manifestations and fatal outcome of falciparum malaria in European patients: observations from TropNetEurop and SIMPID Surveillance Data. *Clin Infect Dis*. 2003 Apr 15;36(8):990-5. doi: 10.1086/374224. Epub 2003 Apr 2. PMID: 12684911.
  28. Checkley AM, Smith A, Smith V, et al. Risk factors for mortality from imported falciparum malaria in the United Kingdom over 20 years: an observational study. *BMJ*. 2012;344:e2116. Published 2012 Mar 27. doi:10.1136/bmj.e2116.
  29. Borges MF de SO, Silva IF da, Koifman R. Histórico social, demográfico e de saúde dos povos indígenas do estado do Acre, Brasil. *Ciênc saúde coletiva* [Internet]. 2020Jun;25(6):2237–46. Available from: <https://doi.org/10.1590/1413-81232020256.12082018>.
  30. BRASIL. Secretaria de Saúde Indígena [online]. Disponível em: [www.gov.br/saude/pt-br/composicao/sesai](http://www.gov.br/saude/pt-br/composicao/sesai) acessado em: 08 de outubro de 2023.
  31. Ministry of Health Brazil, 2016. Health Disease and Attention in Indigenous Territories. Disponível em: [inserir URL]. Acessado em 4 de maio de 2020.
  32. de Aguiar Barros J, Granja F, Pequeno P, Marchesini P, Ferreira da Cruz MF. Gold miners augment malaria transmission in indigenous territories of Roraima state, Brazil. *Malar J*. 2022;21(1):358. doi:10.1186/s12936-022-04381-6.
  33. Ferreira MU, Castro MC. Challenges for malaria elimination in Brazil. *Malar J*. 2016;15(1):284. doi:10.1186/s12936-016-1335-1.
  34. Castro MC, Peterka C. Malaria is increasing in Indigenous and artisanal mining areas in the Brazilian Amazon. *Nat Med*. 2023;29(4):762-764. doi:10.1038/s41591-023-02280-0.

35. Rakislova N, Jordao D, Ismail MR, et al. Accuracy of verbal autopsy, clinical data and minimally invasive autopsy in the evaluation of malaria-specific mortality: an observational study. *BMJ Glob Health*. 2021;6(6):e005218. doi:10.1136/bmjgh-2021-005218.
36. Fonseca MG, Coeli CM, de Fátima de Araújo Lucena F, Veloso VG, Carvalho MS. Accuracy of a probabilistic *record linkage* strategy applied to identify deaths among cases reported to the Brazilian AIDS surveillance database. *Cad Saude Publica*. 2010;26(7):1431-1438. doi:10.1590/s0102-311x2010000700022.
37. Centers for Disease Control and Prevention (CDC). Electronic *record linkage* to identify deaths among persons with AIDS--District of Columbia, 2000-2005. *MMWR Morb Mortal Wkly Rep*. 2008;57(23):631-634.
38. Coelho Neto GC, Chioro A. Afinal, quantos Sistemas de Informação em Saúde de base nacional existem no Brasil? [After all, how many nationwide Health Information Systems are there in Brazil?]. *Cad Saude Publica*. 2021;37(7):e00182119. doi:10.1590/0102-311X00182119.
39. Bronzan RN, McMorrow ML, Kachur SP. Diagnosis of malaria: challenges for clinicians in endemic and non-endemic regions. *Mol Diagn Ther*. 2008;12(5):299-306. doi:10.1007/BF03256295.
40. Adams JH, Mueller I. The Biology of *Plasmodium vivax*. *Cold Spring Harb Perspect Med*. 2017;7(9):a025585. doi:10.1101/cshperspect.a025585.
41. Severe malaria. *Trop Med Int Health*. World Health Organization; 2014;19(s1):7–131. doi:10.1111/tmi.12313\_2.
42. de Laval F, Maugey N, Bonet d'Oléon A, Pommier de Santi V, Ficko C. Increased risk of severe malaria in travellers during the COVID-19 pandemic. *J Travel Med*. 2021;28(6):taab106. doi:10.1093/jtm/taab106.
43. Douglas NM, Poespoprodjo JR, Patriani D, Malloy MJ, Kenangalem E, Sugiarto P, et al. Unsupervised primaquine for the treatment of *Plasmodium vivax* malaria relapses in southern Papua: a hospital-based cohort study. *PLoS Med*. 2017;14:e1002379. doi:10.1371/journal.pmed.1002379.
44. Devine A, Parmiter M, Chu CS, Bancone G, Nosten F, Price RN, et al. Using G6PD tests to enable the safe treatment of *Plasmodium vivax* infections

with primaquine on the Thailand-Myanmar border: a cost-effectiveness analysis. *PLoS Negl Trop Dis.* 2017;11:e0005602. doi:10.1371/journal.pntd.0005602.

45. Bronzan RN, Taylor TE, Mwenechanya J, Tembo M, Kayira K, Bwanaisa L, Njobvu A, Kondowe W, Chalira C, Walsh AL, Phiri A, Wilson LK, Molyneux ME, Graham SM. Bacteremia in Malawian children with severe malaria: prevalence, etiology, HIV coinfection, and outcome. *J Infect Dis.* 2007;195:895–904.
46. Berg A, Patel S, Aukrust P, et al. Increased severity and mortality in adults co-infected with malaria and HIV in Maputo, Mozambique: a prospective cross-sectional study. *PLoS One.* 2014;9(2):e88257. doi:10.1371/journal.pone.0088257.
47. Munyenembe AU, Gausi K, Hiestand J, Mallewa J, Mandala W. The Effect of Frequent Exposure to *P. falciparum*, HIV-Infection, and Other Co-Morbidities on Development of Severe Malaria in Malawian Adults. *Infect Drug Resist.* 2020;13:63-68. doi:10.2147/IDR.S230112.

## 6. CONSIDERAÇÕES FINAIS

Nesta tese, foi descrito o perfil e cenário epidemiológico da malária no território brasileiro ao longo da última década. A investigação do cenário e do comportamento das séries temporais da doença permitiu a compreensão de como se deu o comportamento epidemiológico da malária, com suas mudanças e tendências.

Evidenciou-se que existem dois cenários distintos no Brasil, o cenário da região Amazônica (AMZ) e da região extra-AMZ. Na região extra-AMZ, predominam as notificações de casos importados, provenientes de outros países e da região AMZ brasileira, decorrentes de atividades produtivas ligadas a Viagens e Turismo. Há também a transmissão de casos autóctones, porém, a ocorrência é baixa. Entende-se aqui que: atividades de fortalecimento da vigilância epidemiológica na região extra-AMZ, com detecção precoce e oportuna, poderão evitar a transmissão de casos decorrentes de casos importados e poderão prevenir o desenvolvimento de formas mais graves da doença devido à detecção e diagnóstico tardio. Além disso, a redução de casos de malária na região AMZ influenciará diretamente a ocorrência de casos na região extra-AMZ.

Na região Amazônica, o cenário epidemiológico passou por mudanças ao longo da última década, com momentos de forte redução de casos e, também, o oposto, com um forte aumento de casos e posterior retomada de controle. Há estados que representam desafios crônicos para o controle da malária, como o estado do Amazonas e Roraima. O grande desafio para os próximos anos é o controle da doença em áreas indígenas e de garimpo, devido à complexidade que este controle envolve, como o envolvimento de diferentes setores da administração pública e o respeito pelos aspectos culturais dessas populações. Ainda assim, há estimativas que mostram que o cumprimento das metas intermediárias de redução de casos, para fins de eliminação, poderá ser alcançado nos próximos anos.

Também se evidenciou que o Ministério da Saúde possui grandes limitações e desafios no que tange os seus Sistemas de Informação em Saúde. A ausência de interoperabilidade de seus sistemas dificulta o acompanhamento de casos de malária, dificultando a compreensão dos motivos do desenvolvimento e das formas graves da malária. Apesar das limitações digitais, a utilização de técnicas específicas de vinculação de dados permite a execução deste monitoramento, ainda que os desafios de operacionalização destas técnicas residam na capacidade intelectual dos técnicos da vigilância epidemiológica das diferentes esferas do SUS. A vigilância do óbito por malária será beneficiada com estas técnicas, principalmente pela possibilidade da qualificação das informações dos bancos de dados de mortalidade, que apresentam muitos óbitos por malária com espécies de plasmódios não especificadas.

A operacionalização das técnicas de vinculação de dados, entre SIM e o Sivep-Malária e o Sinan, mostrou fatores que estão associados à ocorrência de óbitos por malária no Brasil, especificamente na região AMZ e na região extra-AMZ.

Os fatores associados ao óbito por malária na região AMZ estão relacionados a atividades da vigilância epidemiológica, como a oportunidade de detecção/tratamento e a forma de detecção (ativa/passiva). Além disso, a parasitemia da doença (acima de três cruces), idades mais vulneráveis (menores de um ano e adultos acima de 40 anos) e populações em áreas indígenas também desempenham um papel importante. Na região extra-AMZ, os óbitos estão associados à infecção por espécies diferentes de *Plasmodium vivax*, com alta parasitemia (acima de três cruces) e casos infectados em outros países diferentes do Brasil (casos importados). Apesar dos fatores identificados, atenção é necessária para os óbitos ocorridos por *Plasmodium vivax* que não é comumente associado a formas graves da doença.

Portanto, tem-se aqui a perspectiva de que esta tese contribuiu para uma compreensão abrangente da malária no contexto brasileiro e fornece maiores evidências que subsidiarão o desenvolvimento de ações e políticas

públicas de saúde com otimização de recursos. O Brasil, ao adotar uma abordagem multidisciplinar e baseada em dados, estará bem direcionado para progredir em direção à eliminação da malária, alinhando-se ao Objetivo de Desenvolvimento Sustentável 3.3 de acabar com as epidemias de doenças transmissíveis até 2030 e, assim, melhorar a saúde de sua população.

## 7. REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, A. C. G. et al. High proportions of asymptomatic and submicroscopic *Plasmodium vivax* infections in a peri-urban area of low transmission in the Brazilian Amazon. **Parasites & Vectors**, v. 11, n. 1, p. 194, 2018. DOI: 10.1186/s13071-018-2787-7.

ALMEIDA, D. et al. Examining the quality of *record linkage* process using nationwide Brazilian administrative databases to build a large birth cohort. **BMC Medical Informatics and Decision Making**, v. 20, n. 1, p. 173, 2020. DOI: 10.1186/s12911-020-01192-0.

BAIA-DA-SILVA, D. C. et al. Current vector control challenges in the fight against malaria in Brazil. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 52, p. e20180542, 2019. DOI: 10.1590/0037-8682-0542-2018.

BANDI, P. et al. Temporal trends in motor vehicle fatalities in the United States, 1968 to 2010 - a joinpoint regression analysis. **Injury Epidemiology**, v. 2, n. 1, p. 4, 17 dez. 2015.

BRASIL 2020. **Guia de tratamento da malária no Brasil**. 2020. Disponível em:  
[https://bvsmis.saude.gov.br/bvs/publicacoes/guia\\_tratamento\\_malaria\\_brasil.pdf](https://bvsmis.saude.gov.br/bvs/publicacoes/guia_tratamento_malaria_brasil.pdf). Acessado em 03 de outubro de 2023.

BRASIL 2021A. **Boletim Especial** - Malária 2021. Bol Epidemiol [Internet]. Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Imunização e Doenças Transmissíveis, Coordenação-Geral de Vigilância de Zoonoses e Doenças de Transmissão Vetorial. Nov. 2021A; 1-118.

BRASIL 2021B. Ministério da Saúde desenvolve plataforma com sistemas de informação de vigilância em saúde. Página de internet. Ago. 2021B. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/noticias/ministerio->

da-saude-desenvolve-plataforma-com-sistemas-de-informacao-de-vigilancia-em-saude. Acessado em 05-03-2022.

BRASIL 2022A. **Panorama epidemiológico da malária em 2021: buscando o caminho para a eliminação da malária no Brasil** – Boletim epidemiológico. Coordenação-Geral de Vigilância de Zoonoses e Doenças de Transmissão Vetorial do Departamento de Imunização e Doenças Transmissíveis da Secretaria de Vigilância em Saúde. 2022A. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/m/malaria/situacao-epidemiologica-da-malaria/boletins-epidemiologicos-de-malaria/boletim-epidemiologico-vol-53-no17-2022-panorama-epidemiologico-da-malaria-em-2021-buscando-o-caminho-para-a-eliminacao-da-malaria-no-brasil/view>. Acessado em 03 de outubro de 2023.

BRASIL 2022B. **Guia de Vigilância em Saúde**: Secretaria de Vigilância em Saúde. Departamento de Articulação Estratégica de Vigilância em Saúde. Coordenação-Geral de Desenvolvimento da Epidemiologia em Serviços. Vol. único 5ª. Ed. – Brasília: Ministério da Saúde, 2022B.

BRASIL 2022C. **Elimina Malária Brasil**: Plano Nacional de Eliminação da Malária. 1st ed. Brasília: MS, 2022C. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/svsa/malaria/politicas-de-saude/elimina-malaria-brasil-plano-nacional-de-eliminacao-da-malaria>. Acessado em 03 de outubro de 2023.

BRASIL. LEI Nº 13.709, DE 14 DE AGOSTO DE 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). 2018. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/L13709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm)

BRASIL. PORTARIA GM/MS Nº 217, DE 1º DE MARÇO DE 2023. DIÁRIO OFICIAL DA UNIÃO. Edição: 42 | Seção: 1 | Página: 63. 2023B. Disponível em: <http://vigilancia.saude.mg.gov.br/index.php/download/portaria-gm-ms-no->

217-de-marco-de-2023-lista-nacional-de-notificacao-compulsoria/?wpdmdl=16385. Acessado em 06 de outubro de 2023.

BRASIL. PORTARIA Nº 1932, DE 09 DE OUTUBRO DE 2003 - Institui o Programa Nacional de Prevenção e Controle da Malária, o Comitê Técnico de Acompanhamento e Assessoramento do Programa e dá outras providências. Disponível em: [https://bvsms.saude.gov.br/bvs/saudelegis/gm/2003/prt1932\\_09\\_10\\_2003.html](https://bvsms.saude.gov.br/bvs/saudelegis/gm/2003/prt1932_09_10_2003.html). Acessado em 03 de outubro de 2023.

BRASIL. Sistema de Informação do Programa de Controle da Malária (SISMAL): relatório resumo epidemiológico, 2001. Brasília: Funasa, 2002. (Mimeo).

BRASIL. Tabnet-DATASUS, 2023A. Disponível em: <https://datasus.saude.gov.br/informacoes-de-saude-tabnet/>. Acesso em: 06 de outubro de 2023.

CAMARGO-JR, K. R.; COELI, C. M. Reclink: aplicativo para o relacionamento de bases de dados, implementando o método probabilistic *record linkage*. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 16, n. 2, p. 439-447, abr-jun, 2000.

CHURCHES, T.; CHRISTEN, P.; LIM, K.; ZHU, J. X. Preparation of name and address data for *record linkage* using hidden Markov models. **BMC Medical Informatics and Decision Making**, v. 2, p. 9, 2002. DOI: 10.1186/1472-6947-2-9.

COELHO NETO, G. C.; CHIORO, A. A. Afinal, quantos Sistemas de Informação em Saúde de base nacional existem no Brasil?. **Cadernos de Saúde Pública**, v. 37, n. 7, p. e00182119, 2021.

COURA, J. R.; SUÁREZ-MUTIS, M.; LADEIA-ANDRADE, S. A new challenge for malaria control in Brazil: asymptomatic Plasmodium infection - a review.

**Memórias do Instituto Oswaldo Cruz**, v. 101, n. 3, p. 229-237, maio de 2006.  
DOI: 10.1590/S0074-02762006000300001.

COX, F. E. História da parasitologia humana. *Clinical Microbiology Reviews*, v. 15, n. 4, p. 595-612, outubro de 2002. DOI: 10.1128/cmr.15.4.595-612.2002. Errata em: **Clinical Microbiology Reviews**, v. 16, n. 1, p. 174, janeiro de 2003.

DERESSA, W.; FANTAHUN, M; ALI, A. Malaria-related mortality based on verbal autopsy in an area of low endemicity in a predominantly rural population in Ethiopia. **Malaria Journal**, v. 6, p. 128, 21 set. 2007. DOI: 10.1186/1475-2875-6-128.

FIKRY, M.; INOUE, S. Optimizing Forecasted Activity Notifications with Reinforcement Learning. **Sensors** 2023, 23, 6510.  
<https://doi.org/10.3390/s23146510>

FLANNERY, E. L.; CHATTERJEE, A. K.; WINZELER, E. A. Antimalarial Drug Discovery: Approaches and Progress towards New Medicines. **Nature Reviews Microbiology**, v. 11, n. 12, p. 849-862, dezembro de 2013. DOI: 10.1038/nrmicro3138.

GARCIA, K. K. S. et al. Assessing the impact of the "malaria supporters project" intervention to malaria control in the Brazilian Amazon: an interrupted time-series analysis. **Malaria Journal**, v. 22, n. 1, p. 275, 2023. DOI: 10.1186/s12936-023-04706-z.

GARCIA, K. K. S.; MIRANDA, C. B.; SOUSA, F. N. E. F. Procedures for health data linkage: applications in health surveillance. **Epidemiologia e Serviços de Saúde**, v. 31, n. 3, e20211272, 2022. DOI: 10.1590/S2237-96222022000300004.

GRANNIS, S.J.; OVERHAGE, J.M.; HUI, S; MCDONALD, C.J. Analysis of a probabilistic *record linkage* technique without human review. **AMIA Annu Symp Proc**. 2003:259-63.

HARPER, K.; ARMELAGOS, G. The changing disease-scape in the third epidemiological transition. **International Journal of Environmental Research and Public Health**, v. 7, n. 2, p. 675-697, fev. 2010. DOI: 10.3390/ijerph7020675.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Amazônia legal. Disponível em: <https://www.ibge.gov.br/geociencias/cartas-e-mapas/mapas-regionais/15819-amazonia-legal.html?=&t=o-que-e>. Acesso em: 06 de outubro de 2023.

KHENG, S. et al. Tolerability and safety of weekly primaquine against relapse of *Plasmodium vivax* in Cambodians with glucose-6-phosphate dehydrogenase deficiency. **BMC Medicine**, v. 13, p. 203, 2015. DOI: 10.1186/s12916-015-0441-1.

KIM, H. J.; FAY, M. P.; FEUER, E. J.; MIDTHUNE, D. N. Permutation tests for jointpoint regression with applications to cancer rates. **Statistics in Medicine**, v. 19, n. 3, p. 335-351, 15 fev. 2000.

KOEHLER, A. B.; SNYDER, R. D.; ORD, J. K. Forecasting models and prediction intervals for the multiplicative Holt-Winters method. **International Journal of Forecasting**, v. 17, n. 2, p. 269-286, abr. 2001.

LAPORTA, G. Z. et al. Reaching the malaria elimination goal in Brazil: a spatial analysis and time-series study. **Infectious Diseases of Poverty**, v. 11, n. 1, p. 39, 5 abr. 2022. DOI: 10.1186/s40249-022-00945-5.

LIAO, H.; XU, Z. Approaches to manage hesitant fuzzy linguistic information based on the cosine distance and similarity measures for HFLTSS and their application in qualitative decision making. **Expert Systems with Applications**, v. 42, n. 12, p. 5328-5336, 2015.

LIMA, M. V. M.; LAPORTA, G. Z.. Avaliação de modelos de predição para ocorrência de malária no estado do Amapá, 1997-2016: um estudo ecológico. **Epidemiologia e Serviços de Saúde**, v. 30, n. 1, p. e2020080, 2021.

LIMA, S. V. M. A. et al. Quality of tuberculosis information systems after *record linkage*. **Revista Brasileira de Enfermagem**, v. 73, n. supl 5, e20200536, 2020. DOI: 10.1590/0034-7167-2020-0536.

LOPES, G. Anopheles gambiae no Brasil: antecedentes para um "alastramento silencioso", 1930-1932. **História, Ciências, Saúde-Manguinhos**, v. 26, n. 3, p. 823-839, 2019.

LÜTHI, B.; SCHLAGENHAUF, P. Risk factors associated with malaria deaths in travelers: a literature review. **Travel Medicine and Infectious Disease**, v. 13, n. 1, p. 48-60, 2015. DOI: 10.1016/j.tmaid.2014.04.014.

MALTA, D. C.; DUARTE, E. C. Causas de mortes evitáveis por ações efetivas dos serviços de saúde: uma revisão da literatura. **Ciência & Saúde Coletiva**, v. 12, n. 3, p. 765-776, junho de 2007.

MARQUES, A. C.; CARDENAS, H. Combate à Malária no Brasil: evolução, situação atual e perspectivas. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 27, p. 91-108, 1994.

MARQUES, M. M. et al. Plasmodium vivax chloroquine resistance and anemia in the western Brazilian Amazon. **Antimicrobial Agents and Chemotherapy**, v. 58, n. 1, p. 342-347, 2014. DOI: 10.1128/AAC.02279-12.

MONTEIRO, W. M. et al. G6PD deficiency in Latin America: systematic review on prevalence and variants. **Memórias do Instituto Oswaldo Cruz**, v. 109, n. 5, p. 553-568, 2014. DOI: 10.1590/0074-0276140123.

MURTA, F. L. G. et al. Perceptions about malaria among Brazilian gold miners in an Amazonian border area: perspectives for malaria elimination strategies. **Malaria Journal**, v. 20, n. 1, p. 286, 2021. DOI: 10.1186/s12936-021-03820-0.

NEKKAB, N. et al. Estimated impact of tafenoquine for Plasmodium vivax control and elimination in Brazil: A modeling study. **PLoS Medicine**, v. 18, n. 4, e1003535, 2021. DOI: 10.1371/journal.pmed.1003535.

OLIVEIRA, G. P. de et al. Accuracy of probabilistic and deterministic *record linkage*: the case of tuberculosis. **Revista De Saúde Pública**, v. 50, p. 49, 2016. DOI: 10.1590/S1518-8787.2016050006327.

OLIVEIRA, G. P. et al. Accuracy of probabilistic and deterministic *record linkage*: the case of tuberculosis. **Revista de Saúde Pública**, v. 50, p. 49, 2016. DOI: 10.1590/S1518-8787.2016050006327.

OLIVEIRA-FERREIRA, J. et al. Malária no Brasil: uma visão geral. **Malaria Journal**, v. 9, p. 115, 30 de abril de 2010. DOI: 10.1186/1475-2875-9-115.

OLSON, S. H. et al. Deforestation and malaria in Mâncio Lima County, Brazil. **Emerging Infectious Diseases**, v. 16, p. 1108-1115, 2010.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). Estratégia Técnica Mundial para o Paludismo 2016–2030. Genebra, Suíça. 2015A.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). **Global Malaria Programme - WHO malaria terminology**. Genebra, Suíça. 2021B.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). Ministerial conference on Malária, Amsterdam. 1992, 67: 349- 50.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). **WHO malaria terminology**. Genebra, Switzerland: WHO malaria terminology; 2019.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). **World malaria report 2021**. Genebra: World Health Organization; 2021A.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). **Guidelines for the treatment of malaria**. Third edition. Genebra, Suíça. 2015B.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). **World malaria report 2022**. Genebra: World Health Organization; 2022.

PEIXOTO, H. M. et al. Cost-effectiveness analysis of rapid diagnostic tests for G6PD deficiency in patients with Plasmodium vivax malaria in the Brazilian Amazon. **Malaria Journal**, v. 15, p. 82, 2016. DOI: 10.1186/s12936-016-1140-x.

PINA-COSTA, A. Vigilância da malária na região extra-amazônica: descrição epidemiológica e clínico-laboratorial dos casos atendidos em uma unidade sentinela. 2009. 106 f. Dissertação (Mestrado em Pesquisa Clínica em Doenças Infecciosas) - Instituto Nacional de Infectologia Evandro Chagas, Fundação Oswaldo Cruz, Rio de Janeiro, 2009. Disponível em: <https://www.arca.fiocruz.br/handle/icict/27874>.

PINHEIRO, R. S.; CAMARGO-JR. K. R.; COELI, C. M. Relacionamento de bases de dados em saúde. **Cadernos de Saúde Coletiva**, Rio de Janeiro, v. 14, n. 2, p. 195-196, 2006.

RAVIGLIONE, M.; MAHER, D. Ending infectious diseases in the era of the Sustainable Development Goals. **Porto Biomedical Journal**, v. 2, n. 5, p. 140-142, 2017. DOI: 10.1016/j.pbj.2017.08.001.

REIS-SANTOS, B. Health Information Systems: how much progress are we making? **Epidemiologia e Serviços de Saúde**, v. 32, n. 2, p. e2022433, 2023.

RICH, S. M. et al. The origin of malignant malaria. **Proceedings of the National Academy of Sciences**, v. 106, n. 35, p. 14902-14907, setembro de 2009. DOI: 10.1073/pnas.0907740106.

SARKAR, J. et al. Risk factors for malaria deaths in Jalpaiguri district, West Bengal, India: evidence for further action. **Malaria Journal**, v. 8, p. 133, 2009. DOI: 10.1186/1475-2875-8-133.

SEVERINI, C.; MENEGON, M. Resistance to antimalarial drugs: An endless world war against Plasmodium that we risk losing. **Journal of Global Antimicrobial Resistance**, v. 3, p. 58-63, 2015. DOI: 10.1016/j.jgar.2015.02.002.

SIQUEIRA, A. M. et al. Malária na Atenção Primária à Saúde. Belo Horizonte: Nescon/UFMG, 2020. 162 p. Disponível em: [https://www.nescon.medicina.ufmg.br/biblioteca/imagem/Malaria-Atencao-Primaria-Saude\\_05\\_11\\_2020%20\\_ATUALIZADO.pdf](https://www.nescon.medicina.ufmg.br/biblioteca/imagem/Malaria-Atencao-Primaria-Saude_05_11_2020%20_ATUALIZADO.pdf).

SWAPNAREKHA, H. et al. Multiplicative Holt-Winters Model for Trend Analysis and Forecasting of COVID-19 Spread in India. **SN Computer Science**, v. 2, n. 5, p. 416, 16 set. 2021.

TALAPKO, J. et al. Malaria: The Past and the Present. **Microorganisms**, v. 7, n. 6, p. 179, 2019. DOI: 10.3390/microorganisms7060179.

TAUIL, P. L. A malária no Brasil. **Cadernos de Saúde Pública**, RJ, v. 1, n. 1, p. 71-111, jan/mar, 1985.

TAUIL, P. L. **Malária no Brasil**: Epidemiologia e Controle. In: Saúde Brasil 2009: Uma Análise da Situação de Saúde e da Agenda Nacional e Internacional de Prioridades em Saúde. Brasília: Ministério da Saúde, 2009.


ULLRICH, A. et al. Impact of the COVID-19 pandemic and associated non-pharmaceutical interventions on other notifiable infectious diseases in Germany: An analysis of national surveillance data during week 1-2016 - week 32-2020. *Lancet Regional Health – Europe*, v. 6, 100103, 2021. DOI: 10.1016/j.lanepe.2021.100103.

WADI, I. et al. Malaria transmission-blocking drugs: implications and future perspectives. **Future Medicinal Chemistry**, v. 12, n. 11, p. 1071-1101, 2020. DOI: 10.4155/fmc-2020-0026.

ZHANG, S.; HU, Y.; BIAN, G. Research on string similarity algorithm based on Levenshtein Distance. In: 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). IEEE, 2017. p. 2247-2251.

## 8. ANEXO I

### Parecer consubstanciado pelo Comitê de Ética em Pesquisa da Faculdade de Medicina da Universidade de Brasília

<p>FACULDADE DE MEDICINA DA UNIVERSIDADE DE BRASÍLIA - UNB</p> 
<b>PARECER CONSUBSTANCIADO DO CEP</b>
<b>DADOS DO PROJETO DE PESQUISA</b>
<b>Título da Pesquisa:</b> Distribuição espaço-temporal de óbitos por malária com estudo de seus fatores de risco e causas associadas no Brasil de 2011 a 2020.
<b>Pesquisador:</b> KLAUSS KLEYDMANN SABINO GARCIA
<b>Área Temática:</b>
<b>Versão:</b> 1
<b>CAAE:</b> 51246121.0.0000.5558
<b>Instituição Proponente:</b> Núcleo de Medicina Tropical
<b>Patrocinador Principal:</b> Financiamento Próprio
<b>DADOS DO PARECER</b>
<b>Número do Parecer:</b> 5.008.210
<b>Apresentação do Projeto:</b>
Estudo do tipo caso-controle com componente ecológico transversal geoespacial com utilização de dados secundários. A população de estudo será de indivíduos registrados no Sistema de Informação de Mortalidade (SIM) classificados como óbitos por malária e indivíduos no Sistema de Informações Hospitalares (SIH) internados por malária que tenham evoluído a óbito. Serão analisados óbitos ocorridos em todo o território brasileiro, incluindo região da Amazônia legal que é composta pelos estados do Acre, Amapá, Amazonas, Mato Grosso, Maranhão, Pará, Rondônia, Roraima e Tocantins e demais estados e Distrito Federal na região extra-amazônica. O período análise será de 2011 a 2020. Para manejo, tratamento e análise dos dados serão utilizados os seguintes softwares: R versão 4.0.3, Stata versão 12, Tableau versão 2020.2, Microsoft Excel 2016, SatScan versão 9.7 e Qgis versão 2.18 e versão 3.16.
<b>Objetivo da Pesquisa:</b>
Identificar fatores prognósticos associados ao risco de morte por malária e seu comportamento espaço-temporal no Brasil de 2011 a 2020.
<b>Endereço:</b> Universidade de Brasília, Campus Universitário Darcy Ribeiro - Faculdade de Medicina <b>Bairro:</b> Asa Norte <b>CEP:</b> 70.910-900 <b>UF:</b> DF <b>Município:</b> BRASILIA <b>Telefone:</b> (61)3107-1918 <b>E-mail:</b> cepfm@unb.br

Continuação do Parecer: 5.008.210

**Avaliação dos Riscos e Benefícios:**

Segundo o pesquisador:

**Riscos:**

Este estudo empregará técnicas e métodos retrospectivos com utilização de informações já coletadas pelo Estado Brasileiro e disponíveis em sistemas de informação. Dessa forma, a pesquisa oferta riscos mínimos à população de estudo e ao considerar o disposto na Resolução 466/2012 e na Lei Geral de Proteção de Dados Pessoais - LGPD (LEI Nº 13.709, DE 14 DE AGOSTO DE 2018) esta pesquisa dispõe de riscos relacionados à garantia do sigilo e a confidencialidade dos dados pessoais dos indivíduos participantes da pesquisa, preservando a privacidade de dados confidenciais. Para assegurar o sigilo e confidencialidade das informações a serem utilizadas os bancos de dados serão solicitados por meio de ofícios, enviados à Secretaria de Vigilância em Saúde do Ministério da Saúde, acompanhados de Termos de responsabilidade de utilização dos dados assinados pelo pesquisador principal da pesquisa. Além disso, as informações sensíveis que permitem identificação do indivíduo serão utilizadas para união de diferentes bancos de dados e este procedimento será feito de acordo com critérios pré-estabelecidos de segurança da informação do Ministério da Saúde. Após a etapa de união dos bancos as informações sensíveis não serão mais utilizadas.

**Benefícios:**

Esta pesquisa não apresenta benefícios diretos individuais a cidadãos brasileiros ou estrangeiros. A realização desta pesquisa apresenta benefícios indiretos pois fornecerá evidências às organizações públicas do Brasil que subsidiarão o desenvolvimento de políticas públicas em saúde que gerarão melhorias à saúde da população brasileira, principalmente à saúde da população acometida por malária.

**Comentários e Considerações sobre a Pesquisa:**

Trata-se de pesquisa de Doutorado do Programa de Pós-Graduação em Medicina Tropical no Núcleo de Medicina Tropical da Universidade de Brasília.

Endereço: Universidade de Brasília, Campus Universitário Darcy Ribeiro - Faculdade de Medicina  
Bairro: Asa Norte CEP: 70.910-900  
UF: DF Município: BRASILIA  
Telefone: (61)3107-1918 E-mail: cepfm@unb.br

Continuação do Parecer: 5.008.210

**Considerações sobre os Termos de apresentação obrigatória:**

Apresentou todos os termos obrigatórios, solicitou dispensa do TCLE.

TCLE: Solicitou dispensa do Termo de consentimento livre e esclarecido, alegando que tal pedido se justifica uma vez que a pesquisa utilizará dados secundários.

**Recomendações:**

Não existem recomendações.

**Conclusões ou Pendências e Lista de Inadequações:**

O projeto está adequado em relação às exigências deste Comitê de ética, por isso foi aprovado.

**Considerações Finais a critério do CEP:**

Após apreciação na reunião dia 29/09/2021 do colegiado CEP/FM o projeto foi aprovado.OBS: De acordo com a Resolução CNS 466/12, nos inciso II.19 e II.20, cabe ao pesquisador elaborar e apresentar ao CEP os relatórios parciais e final do seu projeto de pesquisa. Bem como a notificação de eventos adversos, de emendas ou modificações no protocolo para apreciação do CEP.

**Este parecer foi elaborado baseado nos documentos abaixo relacionados:**

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1776937.pdf	19/08/2021 15:02:14		Aceito
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1776937.pdf	19/08/2021 15:02:12		Aceito
Orçamento	Planilha_de_Orçamento.pdf	19/08/2021 15:01:52	KLAUSS KLEYDMANN SABINO GARCIA	Aceito
Outros	Carta_de_resposta_a_pendencias.pdf	19/08/2021 12:30:40	KLAUSS KLEYDMANN SABINO GARCIA	Aceito
Outros	Curriculos_Lattes_pesquisador_principal_Klauss_Kleydmann_Sabino_Garcia.pdf	19/08/2021 12:28:17	KLAUSS KLEYDMANN SABINO GARCIA	Aceito
Outros	Curriculos_Lattes_Walter_Massa_Ramalhoso.pdf	19/08/2021 12:27:30	KLAUSS KLEYDMANN SABINO GARCIA	Aceito

Endereço: Universidade de Brasília, Campus Universitário Darcy Ribeiro - Faculdade de Medicina  
Bairro: Asa Norte CEP: 70.910-900  
UF: DF Município: BRASÍLIA  
Telefone: (61)3107-1918 E-mail: cepfm@unb.br

FACULDADE DE MEDICINA DA  
UNIVERSIDADE DE BRASÍLIA -  
UNB



Continuação do Parecer: 5.008.210

Outros	Curriculos_Lattes_Anderson_Coutinho.pdf	19/08/2021 12:27:10	KLAUSS KLEYDMANN SABINO GARCIA	Aceito
Outros	Curriculos_Lattes_AmandaAbrahamo.pdf	19/08/2021 12:26:57	KLAUSS KLEYDMANN SABINO GARCIA	Aceito
Outros	Curriculos_Lattes_Ana_Flavia_de_Morais_Oliveira.pdf	19/08/2021 12:25:53	KLAUSS KLEYDMANN SABINO GARCIA	Aceito
Outros	Carta_de_encaminhamento_de_projeto.pdf	19/08/2021 12:24:06	KLAUSS KLEYDMANN SABINO GARCIA	Aceito
Outros	Resumo_estruturado.pdf	19/08/2021 12:19:44	KLAUSS KLEYDMANN SABINO GARCIA	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	Dispensa_de_TCLE.pdf	19/08/2021 12:08:38	KLAUSS KLEYDMANN SABINO GARCIA	Aceito
Declaração de Pesquisadores	Declaracao_de_pesquisadores_Walter.pdf	19/08/2021 12:05:35	KLAUSS KLEYDMANN SABINO GARCIA	Aceito
Declaração de Pesquisadores	Declaracao_de_pesquisadores_Klauss.pdf	19/08/2021 12:05:14	KLAUSS KLEYDMANN SABINO GARCIA	Aceito
Declaração de Pesquisadores	Declaracao_de_pesquisadores_Anderson.pdf	19/08/2021 12:05:06	KLAUSS KLEYDMANN SABINO GARCIA	Aceito
Declaração de Pesquisadores	Declaracao_de_pesquisadores_Ana.pdf	19/08/2021 12:04:54	KLAUSS KLEYDMANN SABINO GARCIA	Aceito
Declaração de Pesquisadores	Declaracao_de_pesquisadores_Amanda.pdf	19/08/2021 12:04:42	KLAUSS KLEYDMANN SABINO GARCIA	Aceito
Declaração de concordância	Declaracao_de_concordancia.pdf	19/08/2021 12:04:07	KLAUSS KLEYDMANN SABINO GARCIA	Aceito
Cronograma	Cronograma_de_pesquisa.pdf	19/08/2021 12:02:38	KLAUSS KLEYDMANN SABINO GARCIA	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_detalhado_Brochura_Pesquisador.pdf	19/08/2021 12:02:19	KLAUSS KLEYDMANN SABINO GARCIA	Aceito
Folha de Rosto	Folha_de_rosto_KlaussGarcia.pdf	19/08/2021 11:56:44	KLAUSS KLEYDMANN SABINO GARCIA	Aceito

Endereço: Universidade de Brasília, Campus Universitário Darcy Ribeiro - Faculdade de Medicina  
Bairro: Asa Norte CEP: 70.910-900  
UF: DF Município: BRASÍLIA  
Telefone: (61)3107-1918 E-mail: cepfm@unb.br

FACULDADE DE MEDICINA DA  
UNIVERSIDADE DE BRASÍLIA -  
UNB



Continuação do Parecer: 5.008.210

**Situação do Parecer:**

Aprovado

**Necessita Apreciação da CONEP:**

Não

BRASILIA, 30 de Setembro de 2021

---

**Assinado por:**

**Antônio Carlos Rodrigues da Cunha**  
(Coordenador(a))

**Endereço:** Universidade de Brasília, Campus Universitário Darcy Ribeiro - Faculdade de Medicina  
**Bairro:** Asa Norte **CEP:** 70.910-900  
**UF:** DF **Município:** BRASILIA  
**Telefone:** (61)3107-1918 **E-mail:** ceptm@unb.br

## 9. ANEXO II

### Ofício de Solicitação de dados à Coordenação-Geral de Vigilância de Zoonoses e Doenças de Transmissão Vetorial, Ministério da Saúde.

#### OFÍCIO

Brasília, 04 de outubro de 2021.

Gabinete da Secretaria de Vigilância em Saúde  
Ministério da Saúde - MS

Ao,  
Senhor Secretário, de Vigilância em Saúde Arnaldo Correia de Medeiros e,  
Coordenador-Geral de Informações e Análises Epidemiológicas,


**Assunto:** Solicitação de acesso a bancos de dados nominais do Sistema de Informação de Mortalidade (SIM) para óbitos Malária (B50 a B54) no Brasil.

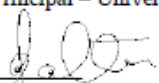
Prezados Senhores,

Ao cumprimentar Vossa Senhoria, venho por meio deste solicitar a cessão dos bancos nominais referidos no termo em anexo para execução do projeto de pesquisa “Distribuição espaço-temporal de óbitos por malária com estudo de seus fatores de risco e causas associadas no Brasil de 2011 a 2020”, o qual compõe minha tese e pesquisa de doutorado, sob orientação de Walter Massa Ramalho professor adjunto da Universidade de Brasília, pelo Programa de Pós-Graduação em Medicina Tropical (PPGMT) do Núcleo de Medicina Tropical da Universidade de Brasília (NMT/UnB).

Esperamos que o desenvolvimento desta pesquisa e seus resultados colaborem para a maior compreensão de fatores relacionados à malária e à ocorrência de óbitos pela doença, assim, subsidiando o desenvolvimento de políticas públicas de saúde que visem controlar e prevenir estes eventos em Saúde.

Atenciosamente,

  
Klauss Kleydmann Sabino Garcia  
Pesquisador Principal – Universidade de Brasília

  
Walter Massa Ramalho  
Orientador e Professor Adjunto – UnB

Núcleo de Medicina Tropical – Faculdade de Medicina  
Campus Universitário Darcy Ribeiro, Brasília-DF | CEP 70910-900

## 10. ANEXO III

### Ofício de Solicitação de dados à Coordenação-Geral de Informações e Análises Epidemiológica, Ministério da Saúde.

#### OFÍCIO

Brasília, 04 de outubro de 2021.

Gabinete da Secretaria de Vigilância em Saúde  
Ministério da Saúde - MS

Ao,  
Senhor Secretário, de Vigilância em Saúde Arnaldo Correia de Medeiros e,  
Coordenador-Geral de Informações e Análises Epidemiológicas,

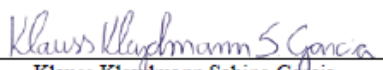
**Assunto:** Solicitação de acesso a bancos de dados nominais do Sistema de Informação de Mortalidade (SIM) para óbitos Malária (B50 a B54) no Brasil.

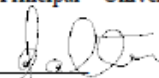
Prezados Senhores,

Ao cumprimentar Vossa Senhoria, venho por meio deste solicitar a cessão dos bancos nominais referidos no termo em anexo para execução do projeto de pesquisa “Distribuição espaço-temporal de óbitos por malária com estudo de seus fatores de risco e causas associadas no Brasil de 2011 a 2020”, o qual compõe minha tese e pesquisa de doutorado, sob orientação de Walter Massa Ramalho professor adjunto da Universidade de Brasília, pelo Programa de Pós-Graduação em Medicina Tropical (PPGMT) do Núcleo de Medicina Tropical da Universidade de Brasília (NMT/UnB).

Esperamos que o desenvolvimento desta pesquisa e seus resultados colaborem para a maior compreensão de fatores relacionados à malária e à ocorrência de óbitos pela doença, assim, subsidiando o desenvolvimento de políticas públicas de saúde que visem controlar e prevenir estes eventos em Saúde.

Atenciosamente,

  
Klauss Kleydmann Sabino Garcia  
Pesquisador Principal – Universidade de Brasília

  
Walter Massa Ramalho  
Orientador e Professor Adjunto – UnB

Núcleo de Medicina Tropical – Faculdade de Medicina  
Campus Universitário Darcy Ribeiro, Brasília-DF | CEP 70910-900

## 11. ANEXO IV

### Artigo publicado – Objetivo 1

Garcia et al. *Malaria Journal* (2022) 21:157  
<https://doi.org/10.1186/s12936-022-04162-1>

Malaria Journal

RESEARCH

Open Access



# Malaria time series in the extra-Amazon region of Brazil: epidemiological scenario and a two-year prediction model

Klauss Kleydmann Sabino Garcia<sup>1\*</sup>, Amanda Amaral Abrahão<sup>2</sup>, Ana Flávia de Moraes Oliveira<sup>3</sup>, Karina Medeiros de Deus Henriques<sup>4</sup>, Anielle de Pina-Costa<sup>5</sup>, André Machado Siqueira<sup>4</sup> and Walter Massa Ramalho<sup>1</sup>

## Abstract

**Background:** In Brazil, malaria is caused mainly by the *Plasmodium vivax* and *Plasmodium falciparum* species. Its transmission occurs in endemic and non-endemic areas. Malaria geography in Brazil has retracted and is now concentrated in the North region. The Brazilian Amazon region accounts for 99% of Brazil's cases. Brazil's extra-Amazon region has a high frequency of imported cases and in 2019 presented a mortality rate 123 times higher than the Amazon region. Extra-Amazon cases present risks of reintroduction. This study aims to characterize the epidemiological scenario for malaria in the extra-Amazon region of Brazil from 2011 to 2020 with a two-year forecast.

**Methods:** Time-series study with description of malaria cases and deaths registered in Brazilian extra-Amazon region from 2011 to 2020. Public data from the Notifiable Diseases Information System (Sinan) and the Mortality Information System (SIM) were used. Descriptive analysis, incidence, and notification rates were calculated. Flow charts analysed the flux between Places of Probable Infection (PI) and places of notification. The prediction model utilized a multiplicative Holt-winters model for trend and seasonality components.

**Results:** A total of 6849 cases were registered. Cases were predominantly white males with 9 to 11 years of education, mostly between 30 and 39 years old. Imported cases accounted for 78.9% of cases. Most frequent occupations for imported cases are related to travelling and tourism activities. Among autochthonous cases, there is a higher frequency of agriculture and domestic economic activities. In the period there were 118 deaths due to malaria, of which 34.7% were caused by *P. falciparum* infections and 48.3% were not specified. The most intense flows of imported cases are from Amazonas and Rondônia to São Paulo, Rio de Janeiro, and Paraná. The prediction estimates around 611 cases for each of the following two years.

**Conclusion:** The time series allows a vast epidemiological visualization with a short-term prediction analysis that supports public health planning. Government actions need to be better directed in the extra-Amazon region so the objective of eliminating malaria in Brazil is achieved. Carrying out quality assessments for information systems and qualifying personnel is advisable. Malaria outside the Amazon region is mainly due to imported cases and delay in diagnosis is associated with a higher fatality rate. Better strategies to diagnose and treat suspected cases can lead to lower risk of deaths and local outbreaks that will be important for achieving malaria elimination in Brazil.

\*Correspondence: [kleydmann25@gmail.com](mailto:kleydmann25@gmail.com)

<sup>1</sup>Tropical Medicine Center, University of Brasília, Darcy Ribeiro University Campus, Brasília, Brazil  
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Keywords:** Malaria, Malaria Case, Epidemiology, Public Health, Control; Elimination, Extra-Amazon

## Background

Malaria is an acute febrile infectious disease, caused by a protozoan of the genus *Plasmodium* that infects humans through female *Anopheles* mosquito bites. It is common in tropical and developing countries and is considered a neglected tropical disease by the World Health Organization (WHO). Malaria affects about 212 million people worldwide and caused more than 409,000 deaths in 2019, with mortality rates ranging from 0.2 to 2.2 worldwide [1, 2].

In Brazil, the main *Plasmodium* species are *Plasmodium vivax* and *Plasmodium falciparum*, but there is also the transmission of *Plasmodium malariae*, although its occurrence is low compared to that of *P. vivax* and *P. falciparum*. Transmission by *P. ovale* has not been reported in Brazil in previous literature. Malaria transmission in endemic and non-endemic areas relates to the particular *Anopheles* vector species, common in areas of the Amazon Forest and the Atlantic Forest [3–6].

Historically, malaria in Brazil has undergone changes in its geographic distribution. Its occurrence during the twentieth century was prevalent throughout the country but in the last 50 years it has become concentrated in the Brazilian Amazon region [7, 8]. The Legal Amazon region comprises approximately 99% of all cases in the country. Although the extra-Amazon region accounts for only 1% of the total cases, it is noteworthy that cases in this region presents higher lethality and represent a possible risk for a re-establishment of malaria transmission [9]. From 2011 to 2020, Brazil registered about 1.8 million cases of malaria with 450 deaths [10].

Although the total number of deaths from malaria in the extra-Amazon region is lower than in the legal Amazon region, the relative case fatality was 123 times greater than in the Amazon region in 2019 [3, 11]. Thus, this study aims to describe and characterize the malaria epidemiological scenario and profile in the extra-Amazon region of Brazil over the last 10 years.

## Methods

### Study design

Ecological time-series study with the description of reported cases and deaths from malaria in the Brazil extra-Amazon region from 2011 to 2020. In this study, the following definitions for malaria cases were adopted: Reported case: Case with a positive laboratory result for malaria; new infection (new case): Case with positive laboratory result for malaria excluding positive Cure Slide

Verification (CSV); Autochthonous cases: Classification based on the place of probable infection (municipality or state). Negative cases were not analysed in this study. Malaria deaths were analysed according to the place of residence of the case [12, 13].

### Study site

The study site selected was the Brazilian extra-Amazon region, which includes 17 states (Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe, Bahia, Minas Gerais, Espírito Santo, Rio de Janeiro, São Paulo, Paraná, Santa Catarina, Rio Grande do Sul, Mato Grosso do Sul, Goiás) and the Federal District. It is noteworthy that the Legal Amazon region covers all the states of the North region plus the states of Maranhão, Mato Grosso, and Tocantins [14]. According to the Brazilian Institute of Geography and Statistics (IBGE) population projections, the extra-Amazon region had an average of 176,163,089 inhabitants, per year between 2011 and 2020.

### Data source and analysis

Public data from the Notifiable Diseases Information System (Sinan) were obtained from the Brazilian Ministry of Health (MoH) database portal – DATASUS – in May 2021. MoH provided Mortality data (SIM) through the Integrated Ombudsman and Access to Information Platform of the Brazilian Federal Government in June 2021. For data processing and analysis the following software were used: Qgis Desktop (2.18), Microsoft Excel (2016) and the R software packages read.dbc, tidyverse, lubridate, forecast, and fpp (version 4.0.5) [15].

The analysed variables were: Notification date, Symptom onset date, Treatment start date, Birthdate, Age, Gender, Race/Color, Education, Pregnant woman, Activity in the last 15 days, professional Activities, Type of slide (active, passive detection, CSV), Examination results, Autochthony, Notification locations – Municipality and State (denoted as FU—Federated Units), PI and death location.

### Incidence rates

Malaria incidence rates (total new malaria cases by the PI municipality /Total population of the municipality\*100,000) and malaria notification rates (total malaria cases by the notification municipality /Total population of the municipality \*100,000) were calculated for each year. Incidence rates calculated by the PI consider that those cases are autochthonous. The total new infections number is a result of the total of positive cases in the

period excluding those classified as CSV, which represent recurrences of the infection [12].

The Qgis software plugin "Flowmaps" (version 2.18) was used to analyse flows between PI and the place of notification of new malaria infections. Along with the flow analysis the average incidence rate for the period was described (Average of: Total new infections number by the PI municipality for each year/ Resident population by the municipality for each year  $\times$  100,000 inhabitants).

The R software commands *ts*, *decompose*, *Holt-winters*, and *forecast* were used to predict cases for 2021 and 2022. For the prediction analysis, the attributes of mean, trend, and seasonality of notified malaria cases from 2011 to 2020 were considered. Holt-winters exponential smoothing and prediction methods with a multiplicative seasonality influence in modelling was used, as in the study of Swapnarekha et al. [16]. The Holt-winters seasonal multiplicative model multiplies the series trend by the seasonality, instead of adding one to the other, as it is in the Holt-winters additive modelling. The multiplicative model is better suited to malaria's time series because its seasonality has a higher influence on increasing or decreasing cases number. If the seasonality did not have such a high influence on the time series, the additive model would be a better choice to do the forecast modelling.

The Holt-Winters prediction model evaluates the time series behavior based on the patterns of the three attributes mentioned earlier. Exponential smoothing uses an exponentially weighted moving average (EWMA) to level the time series and provide a higher quality forecast. This model uses exponential smoothing to consider historical values and predicts a likely future value considering 95% confidence intervals [17].

This study did not require approval by ethics committees according to the Brazilian National Health Council resolution n° 466/2012. This work received funding from the Bill & Melinda Gates Foundation (INV-003970) and the Brazilian Ministry of Health (MoH/DECIT/CNPq #443,148/2019–8). All data generated or analysed during this study are included in this published article Additional files 1, 2 and 3.

## Results

From 2011 to 2020, the Brazilian extra-Amazon region registered 6,849 notifications of confirmed malaria cases, of which were 682 (9.9%) cases classified as CSV, and 6,167 (90.1%) were new malaria cases notifications, where 1,640 (26.5%) were notified through active detection.

Cases were predominantly male (78%), white (44.3%), and mixed (38.7%); the distribution by race/colour is similar between males and females. The highest concentration by age group is in 20 to 59 years (84%). The

age group 30 to 39 years presents the highest frequency of cases (26.1%), along with white or brown people (42.8% and 38%, respectively). Among the positive cases, 58 cases were found to have a self-reported age of less than one year, with the median being 37 years (1-quartile: 28; 3-quartile: 49; maximum: 93). As for education, the highest frequencies are in people with 9 to 11 years of schooling (21.7%). For "Education level", 23.6% were filled out as "ignored" (Table 1).

As for occupations, the most frequent activities are related to "travel" and "tourism", which together account for 31.2% of the total number of records for malaria cases in the extra-Amazon region. Women have a higher relative frequency of activities related to "tourism" and "domestic" employees, whereas men have a higher relative frequency of cases in "agriculture", "construction" and "mining" activities (Table 1). Analysing the occupations according to the level of education, the most frequent are "Traveller" and "Tourism", which account for 4.6% and 2.6% when over 12 years of study, respectively.

For autochthonous cases, the most frequent activities are "Agriculture" (264 cases; 26.9%) and "Domestic" (93 cases; 9.5%). Imported cases are more frequent in "Travellers" (1,110 cases; 22.2%) and "Tourism" (630 cases; 12.5%). Information filled as "Other" economical activities account for a total of 25.8% of autochthonous cases and 28.4% of imported cases (Table 2). Information about occupational activities in death notifications shows that the most frequent deaths in the extra-Amazon region are of retired people (16 deaths; 13.7%), housewives (6 deaths; 5.1%) and, managers in commercial activities (6 deaths; 5.1%).

Among the new malaria cases in the period, 912 (15%) records were notified as autochthonous cases, 4,869 (78.9%) as imported and 307 (4.5%) were indeterminate. Malaria due to *P. vivax* accounted for 4,592 cases (67.5%) and *P. falciparum* or mixed malaria cases (*P. falciparum* malaria along with other species) accounted for 2,159 (31.5%) notifications. There is also a total of 39 malaria cases caused by *P. malariae* and 59 cases caused by *P. ovale*, which represent only 1.4% of the total number of confirmed cases over the time series (Table 2).

The time series presented an approximate average of 685 cases per year, with the highest number of cases reported in 2011, at 1,066 cases, with the lowest being 336 cases in 2020. The monthly average was 57 cases (minimum of 10 and maximum of 177). From 2011 to 2020, there were 118 deaths, with a minimum of five deaths in 2017 and a maximum of 18 in 2011 (Fig. 1). Of these, 41 (34.7%) deaths were due to *P. falciparum*, 17 (14.4%) due to *P. vivax*, 3 (2.5%) due to *P. malariae*, and 56 (48.3%) due to unspecified malaria.

**Table 1** Epidemiological profile of malaria cases in the extra-Amazon region of Brazil, 2011 to 2020

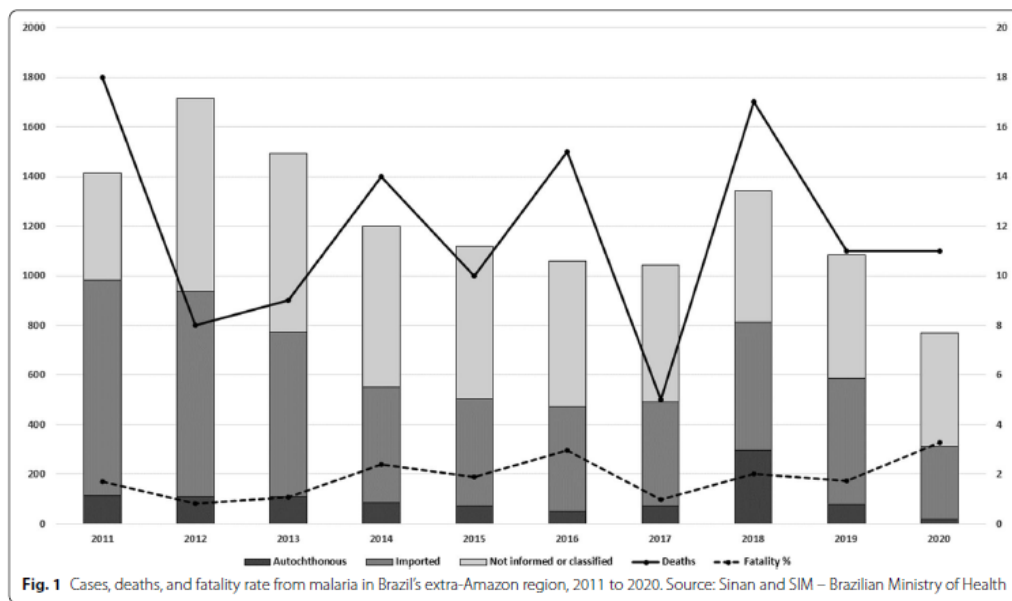
Sex	Male		Women		Total	
	N	%	n	%	N	%
Total	5343	78.01	1506	21.99	6849	100.00
Age groups						
Under 1 year old	45	0.84	14	0.93	59	0.86
01—04	52	0.97	29	1.93	81	1.18
05—09	37	0.69	24	1.59	61	0.89
10—14	85	1.59	54	3.59	139	2.03
15—19	180	3.37	69	4.58	249	3.64
20—29	1096	20.51	357	23.71	1453	21.21
30—39	1461	27.34	324	21.51	1785	26.06
40—49	1116	20.89	265	17.60	1381	20.16
50—59	817	15.29	227	15.07	1044	15.24
60—69	353	6.61	87	5.78	440	6.42
70—79	87	1.63	45	2.99	132	1.93
80 years or more	14	0.26	11	0.73	25	0.37
Total	5343	100	1506	100	6849	100.00
Race/Color						
White	2319	44.35	610	41.16	2,929	43.64
Mixed	2025	38.73	578	39.00	2,603	38.79
Black	470	8.99	158	10.66	628	9.36
Indigenous	17	0.33	16	1.08	33	0.49
Yellow	35	0.67	14	0.94	49	0.73
Ignored	363	6.94	106	7.15	469	6.99
Total	5229	100	1482	100	6711	100.00
Years of study						
None	49	1.0	13	0.9	62	1.0
1 to 4	508	10.4	131	9.4	639	10.2
5 to 8	913	18.7	196	14.1	1109	17.7
9 to 11	1215	24.9	276	19.8	1491	23.8
12 or more	861	17.6	341	24.5	1202	19.2
Not applicable	106	2.2	50	3.6	156	2.5
Ignored	1230.00	25.2	387	27.8	1617.00	25.8
Total	4882	100	1394	100	6276	100
Professional activities						
Traveller	931	19.03	303	20.21	1,23	19.31
Tourism	447	9.14	245	16.34	692	10.83
Agriculture	457	9.34	89	5.94	546	8.54
Panning	443	9.05	40	2.67	483	7.56
Constructor	239	4.88	3	0.20	242	3.79
Mining	192	3.92	39	2.60	231	3.61
Domestic	52	1.06	178	11.87	230	3.60
Hunting/ Fishing	108	2.21	6	0.40	114	1.78
Livestock	68	1.39	5	0.33	73	1.14
Expl. vegetable	57	1.16	5	0.33	62	0.97
Others	141	28.88	397	26.48	181	28.32
Ignored	486	9.93	189	12.61	675	10.56
Total	489	100	15	100	639	100.00

Source: Sinan—Brazilian Ministry of Health

**Table 2** Cases distribution by *Plasmodium* species, professional activity, and autochthony, 2011 to 2020, Brazil's extra-Amazon region

	Autochthonous		Imported		Undetermined		Total	
	n	%	n	%	n	%	n	%
Type of malaria for new infections								
Falci-parum	212	21.1	1805	33.4	96	31.3	2113	31.4
Vivax	785	78.0	3539	65.4	200	65.1	4524	67.3
Ovale	2	0.2	42	0.8	5	1.6	49	0.7
Malariae	8	0.8	25	0.5	6	2.0	39	0.6
Total	1007	100	5411	100	307	100	6725	100
Economic activities								
Traveller	51	5.2	111	22.2	62	21.4	1223	19.5
Tourism	40	4.1	630	12.6	17	5.9	687	10.9
Agriculture	264	26.9	252	5.0	16	5.5	532	8.5
Pannig	37	3.8	420	8.4	13	4.5	470	7.5
Constructor	19	1.9	215	4.3	4	1.4	238	3.8
Domestic	93	9.5	127	2.5	7	2.4	227	3.6
Mining	25	2.5	180	3.6	21	7.2	226	3.6
Hunting/Fishing	15	1.5	95	1.9	3	1.0	113	1.8
Livestock	16	1.6	52	1.0	2	0.7	70	1.1
Vegetal exploration	22	2.2	40	0.8	-	-	62	1.0
Others	253	25.8	1423	28.4	96	33.1	1772	28.2
Ignored	146	14.9	462	9.2	49	16.9	657	10.5
Total	981	100	5006	100	290	100	6277	100

Source: Sinan—Brazilian Ministry of Health

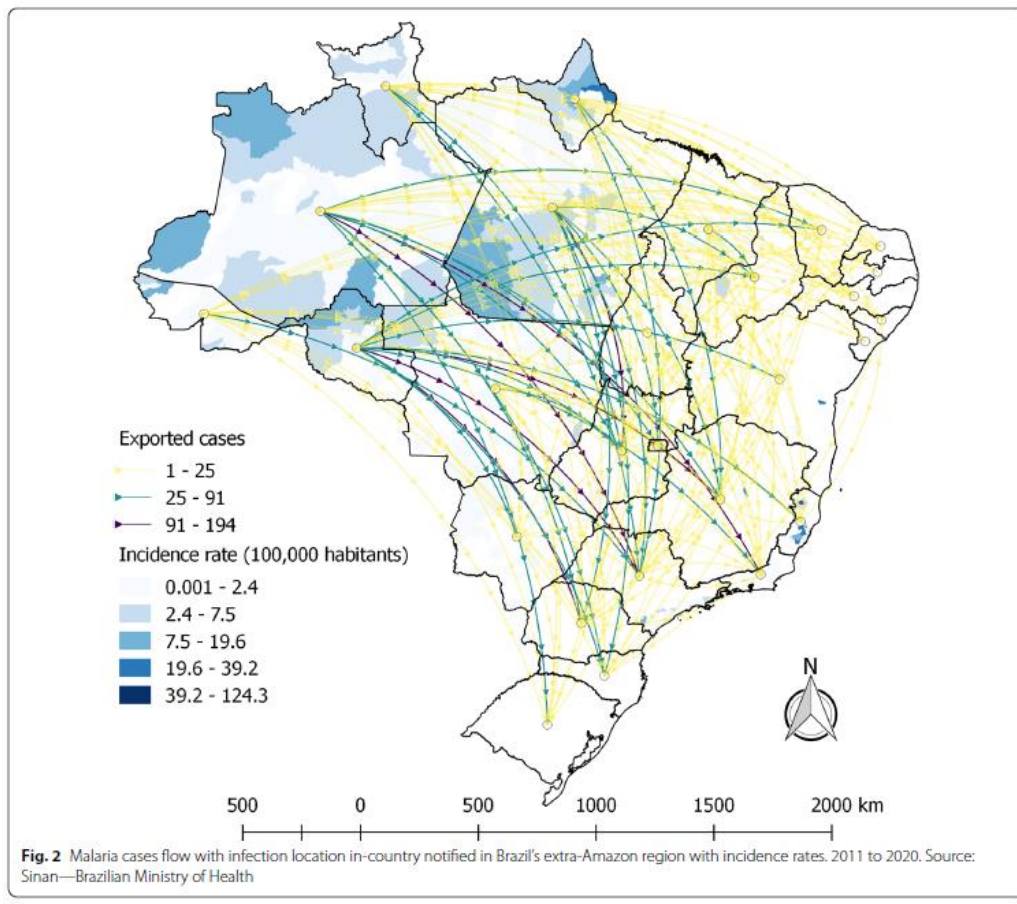


FU most frequent autochthonous cases occurred in Espírito Santo (411 cases; 6%), Bahia (136 cases; 2%) and São Paulo (123 cases; 1.8%). The states that most reported malaria cases were São Paulo (1,405 reported cases; 20.5%), Rio de Janeiro (778; 11.4%), and Minas Gerais (741; 10.8%). For imported cases, most cases are notified in São Paulo (1,214 cases; 22.4%), Minas Gerais (691 cases; 12.8%), and Rio de Janeiro (649 cases; 12%).

Of the 118 deaths in the extra-Amazon region, they were mostly of residents from São Paulo (30 deaths; 25.4%), Minas Gerais (21 deaths; 17.8%), Bahia, and Goiás (9 deaths each; 7.6%). Figure 2 shows that the most intense flows were in the North region, mainly the states of Amazonas and Rondônia to the states of São Paulo, Rio de Janeiro, and Paraná (Fig. 2).

Annual incidence rates for the entire region ranged from 0.16 to 0.60 cases per 100,000 inhabitants, whereas municipal incidence rates ranged from 0.008 (São Paulo/SP in 2020) to 1,243 (Vila Pavão/ES in 2018) cases for every 100,000 inhabitants (average: 11—median: 4.6). The notification rates showed a similar annual variation, ranging from 0.03 (Salvador/BA in 2020) to 1,286.6 (Vila Pavão/ES in 2018) cases reported for every 100,000 inhabitants (mean 6.7—median: 1.6). The extra-Amazon malaria fatality rate from 2011 to 2020 was 1.8, with a minimum of 0.81 in 2012 and a maximum of 3.27 in 2020 (Fig. 1).

As for other imported cases, Brazil notified 1,771 new infections from other countries in the time series. The countries responsible for up to 80% of the total imported



cases are Angola (534–30.2%), South Africa (163–9.2%), Mozambique (158–8.9%), French Guiana (137–7.7%), Guyana (132–7.5%), Nigeria (124–7.0%), Venezuela (116–6.5%) and Equatorial Guinea (60–3.4%).

The prediction model used presented estimates for the months of 2021 and 2022, with an average of 27 cases per month in the first year and 24 monthly cases in 2022. The time series showed a cases reduction trend. The estimates are accompanied by confidence intervals that, when predicted for periods longer than 12 months start to lose precision in their estimates, thus, it is noteworthy that for the year 2021, 324 cases were predicted with a minimum of 16 and a maximum of 47 cases in a single month (Fig. 3). 611 cases were predicted for the following two years.

### Discussion

Considering that Brazil has a goal of eliminating malaria cases by the year 2035 [11] new malaria cases and recurrences were analysed together in the epidemiological profile, even though the database recurrences number are probably underestimated. From the perspective of elimination, recurrence cases must be considered in the time series analysis of the profile, because an untreated recurrence case not treated may become an outbreak source, and it needs to receive malaria treatment. As for the predictive analysis, only new malaria cases were used because 79% of the cases notified in the extra-Amazon region were imported new infections from the Amazon region—as in the study of Machado and collaborators [18]. It is important to say that CSV is underestimated in Sinan because imported cases could move back to their regions and so it is notified in a different information system or another country.

The results highlight that the occurrence of malaria cases in men, as it is in the Amazon region, is higher than in women. Additionally, the cases are made up of mostly adults (20 to 59 years) and have a similar distribution between people of white and brown color/race. The distribution of malaria cases by age group, for the same period, differs from the profile of cases reported in the Amazon region, where the concentration of cases in people under 20 years of age is approximately 46.3% of the total cases. In the extra-Amazon region, the portion of people under 20 years of age does not exceed 9%. As for the distribution of cases by race/colour, the profile is also different because in the Amazon region about 60.8% of cases are brown, 17.6% indigenous and only 6.2% are white (data updated on 06–09-2021) [19, 20].

Considering that most cases reported in the extra-Amazon region are imported cases and that travelling to endemic areas is a known risk factor for contracting malaria [21], this difference in the profile may be related

to socio-economic issues such as years of schooling, economic activity, and income. This indicates that the occurrence of malaria cases in the extra-Amazon and Amazon regions of Brazil relates to social inequalities.

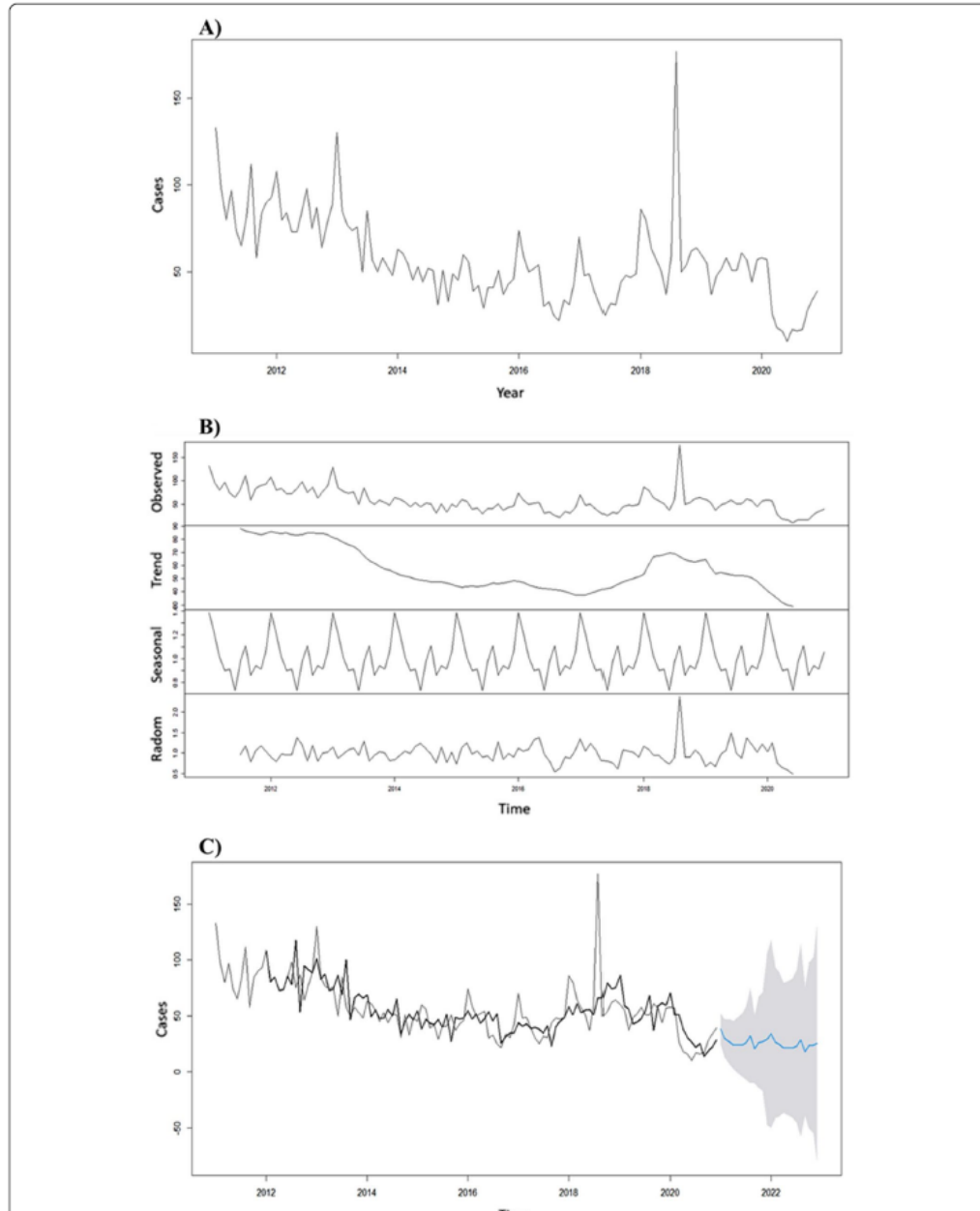
The MoH traditionally presents different strategies for both regions, however, it is suggested that specific internal control goals for the country be worked out for the extra-Amazon region [11, 12, 19].

The notification of autochthonous cases in the extra-Amazon region is a warning to epidemiological surveillance centers as it represents the establishment of community transmission, which is a risk for the reintroduction of malaria in these non-endemic, but still susceptible regions [22]. It is important to highlight that approximately 27.5% of the extra-Amazon cases were notified in the year 2018, which was an atypical year for malaria control in Brazil, as the country was showing constant reductions in malaria rates until 2017. In 2016 Brazil had reported approximately 129,000 cases and in 2017 reported over 194,000 [11].

When analysing the PI and the notification place in the extra-Amazon region according to their professional activity, it became clear that a large number of imported malaria cases are from people with professional activities related to travelling or tourism, with the mining activities also standing out. There is a portion of cases that were classified as indeterminate autochthony which are patients with professional activity also related to travelling [23, 24]. Agriculture activities stand out as the most frequent activity, being the most frequent between autochthonous cases. This may be related to the area that they work or live in, considering that these activities may be agriculture for self-sustaining. The high number of cases with undetermined PI also stood out. It should be noted that epidemiological information is as important as the response to possible outbreaks. It is advisable that specific studies be carried out to determine job-related malaria characteristics in Brazil, given that over 38% of economic activities are blank, ignored, or defined as "Other".

The data has shown two autochthonous cases of malaria by *Plasmodium ovale* in Brazil, which is a parasitic species with no previous records of community transmission in the country [3, 22, 25]. These notifications should be reviewed by the state and municipal health departments to better understand these cases and certify whether or not there was community transmission. These notifications may be system-filling or diagnostic errors because they were isolated cases, one in 2011 infected in the state of Rondônia and reported in Rio de Janeiro and another in 2014 infected and reported in Espírito Santo.

It is important to highlight that 4.4% of the records for the period had no defined autochthony. Therefore,



**Fig. 3** Brazilian extra-Amazon region Malaria time series, trend, seasonality, and randomness attributes with two-year prediction. Source: Sinan—Brazilian Ministry of Health. **A** Brazilian extra-Amazon region malaria time series; **B** Malaria time series components; **C** Malaria cases prediction for 2021 and 2022

epidemiological investigation in the extra-Amazon region needs improvement.

As long as the Amazon region presents a high incidence of autochthonous malaria cases the extra-Amazon case notifications will not decrease. It is necessary to understand that the occurrence of autochthonous and imported cases in the extra-Amazon region follows a seasonal pattern and a trend. Those components' behaviour is related to the malaria epidemiological behavior in the Amazon region [18, 23].

It is critical that specific actions and policies for preventing deaths be directed to the extra-Amazon region, mainly in the states of São Paulo, Minas Gerais, and Goiás, which are the ones with the highest accumulated number of deaths in the last 10 years. Those states may have the highest number of deaths because there have the largest fluxes of travellers and people exposed to infections. Therefore, they are the states with highest number of cases and the number of deaths and fatality rate reflects failures in the identification and opportune treatment of cases. Policies must be directed at identifying suspected cases, testing them and providing proper care. International studies highlight some common risk factors for the occurrence of deaths, such as cases of non-autochthonous malaria, or that come from endemic regions, or that experience therapeutic failures or treatment delay [26, 27].

As shown by Duarte and collaborators [7] it is necessary to implement control and prevention strategies tailored to the realities of each geographic grouping. In addition, it is necessary to assess the quality of death notifications data, as the highest frequency of deaths was for ICD-10 B54 which is "Malaria not specified". The literature [22], as well as results from this study, suggests a high frequency of deaths from *P. falciparum* in the extra-Amazon region. Linkages techniques should be used in official health information systems for better specification of the species related to malaria deaths.

According to the Brazilian Health Ministry online data, the predictions for the 2021 initial 3 months were close to the value registered in Sinan for malaria cases in the extra-Amazon region (predicted: 108; observed: 113) [20]. Thus, the developed prediction allows better planning of specific control strategies for the containment of malaria cases, especially when considering the seasonality component of the disease [23, 28, 29]. The trend towards a reduction in cases in the extra-Amazon region can be interpreted as a result of the success in disease control and prevention actions carried out on the national and local levels. The model used in this study can be replicated on a state level.

The decrease in 2020 may relate to travel restrictions in the face of the COVID-19 pandemic, thus reducing

visits to endemic areas in the North region of the country. This behaviour was previously noted in notifiable diseases in Australia and Taiwan, and this may also have occurred in Brazil [30, 31].

As for the limitations of this study, there is the fact that secondary data were used; the database had some missing information, which is shown by the results tabulations values that do not match each other. This difference is because some notifications do not have all the information on their investigation form filled out. In addition, defining autochthony of death is not possible due to SIM databank limitations. Linkage techniques may be an interesting strategy to present additional and more accurate information. Lastly, occupational activities in the SIM database may not be precise; therefore, further investigation is advised.

## Conclusions

A preventive and early detection measure that could be tested for screening malaria cases is to use rapid tests in bus stations and airports to test travellers presenting symptoms or those very high-risk areas for malaria. Another possible measure is, distributing an information booklet on malaria in these places with the description of the disease and guidance to the health professional on how to act when detecting the main symptoms.

The analysis of time series allows for a better understanding of its epidemiological behavior and, when accompanied by short-term predictions, can serve as tools to support government decision-making.

To achieve malaria elimination the government must invest in qualifying personnel that fill out malaria notifications and diagnose malaria, while also carrying out quality assessments of the Sinan and SIM databases.

Thus, one step to achieve the national elimination goal is the elimination of malaria in the extra-Amazon region of Brazil, preventing cases and deaths from the disease. Although malaria outside the Amazon represents a relatively limited portion of cases, it still accounts for a considerable number of cases and deaths, and also poses a risk for reintroduction illustrated by several local transmission outbreaks. To eliminate malaria from the country it is essential that a more timely and comprehensive surveillance is established in order to reduce these threats through better provision of care to suspected cases.

## Abbreviations

MoH: Ministry of Health; CSV: Cure Slide Verification; FU: Federated Units; ICD: International Classification of Diseases; PI: Place of Probable Infection; SIM: Mortality Information System; Sinan: Notifiable Diseases Information System; WHO: World Health Organization.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12936-022-04162-1>.

**Additional file 1.** A script for software R containing programming used to develop the analysis.

**Additional file 2.** The study database containing malaria cases in Brazil's extra-amazon region for the analyzed period.

**Additional file 3.** The study database containing malaria deaths in Brazil's extra-amazon region for the analyzed period.

## Acknowledgements

Thanks to all authors' contributions and the Nucleus of Tropical Medicine from the University of Brasília for providing knowledge and means to develop the research.

## Author contributions

KKSG performed data analysis, interpretation and wrote the manuscript. KKSG, AAA, AFMO, WMR helped with Project creation and discussion. KMDH, APC, AMS revised the manuscript and helped with the discussion. All authors have read and approved the final manuscript.

## Funding

This work was supported, by the Bill & Melinda Gates Foundation [INV-003970] and the Brazilian MoH/DECIT/CNPq (443148/2019–8). Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission.

## Availability of data and materials

The datasets generated and/or analysed during the current study are available in the OneDrive repository [[https://saudegov-my.sharepoint.com/:f/g/personal/klaus\\_garcia\\_saude\\_gov\\_br/EvBGBH8SpxEoHEYtu7hw1wBFFJ3dyZQUQpQsmdln\\_91cUA7e=UIPAT](https://saudegov-my.sharepoint.com/:f/g/personal/klaus_garcia_saude_gov_br/EvBGBH8SpxEoHEYtu7hw1wBFFJ3dyZQUQpQsmdln_91cUA7e=UIPAT)] and in the public website of MoH (DATASUS), link: <https://datasus.saude.gov.br/transferecia-de-arquivos/>.

## Declarations

### Ethics approval and consent to participate

This study did not require approval by ethics committees in accordance with Brazilian National Health Council resolution nº 446/2011.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Tropical Medicine Center, University of Brasília, Darcy Ribeiro University Campus, Brasília, Brazil. <sup>2</sup>Health Ministry of Brazil, Federal District, Brasília, Brazil. <sup>3</sup>Federal Institute of Education, Science and Technology of Tocantins, IFTO, Araguaína, Brazil. <sup>4</sup>FIOCRUZ, Evandro Chagas National Institute of Infectious Diseases, Av. Brasil, Rio de Janeiro, RJ 4365, Brazil. <sup>5</sup>School of Medicine, Centro Universitário Serra Dos Órgãos (UNIFESO), Teresópolis, RJ, Brazil.

Received: 12 November 2021 Accepted: 13 April 2022

Published online: 31 May 2022

## References

- WHO. World malaria report. 20 years of global progress & challenges. Geneva: World Health Organization; 2020. p. 2020.
- Talapko J, Skrlac I, Alebić T, Jukić M, Včev A. Malaria: the past and the present. *Microorganisms*. 2019;7:179.
- Oliveira-Ferreira J, Lacerda MV, Brasil P, Ladislau JL, Tauil PL, Daniel-Ribeiro CT. Malaria in Brazil: an overview. *Malar J*. 2010;9:115.
- Tadei WP, Dutary TB. Malaria vectors in the Brazilian Amazon: *Anopheles* of the subgenus *Nyssorhynchus*. *Rev Inst Med Trop Sao Paulo*. 2000;42:87–94.
- Hiwat H, Bretas G. Ecology of *Anopheles darlingi* Root with respect to vector importance: a review. *Parasit Vectors*. 2011;4:177.
- Sinka ME, Bangs MJ, Manguin S, Rubio-Palis Y, Chareonviriyaphap T, Coetzee M, et al. A global map of dominant malaria vectors. *Parasit Vectors*. 2012;5:69.
- Duarte EC, Ramalho WM, Tauil PL, Fontes CJ, Pang L. The changing distribution of malaria in the Brazilian Amazon, 2003–2004 and 2008–2009. *Rev Soc Bras Med Trop*. 2014;47:763–9.
- França TCC, dos Santos MG, Figueroa-Villar JD. Malaria: historical aspects and chemotherapy. *Quim Nova*. 2008;31:1271–8.
- Lorenz C, Virginio F, Aguiar BS, Suesdek L, Chiaravalloti-Neto F. Spatial and temporal epidemiology of malaria in extra-Amazonian regions of Brazil. *Malar J*. 2015;14:408.
- Brazilian Ministry of Health, Secretariat of Health Surveillance, Department of Immunization and Communicable Diseases, General Coordination of Surveillance of Zoonoses and Vector-Transmitted Diseases. Epidemiological Bulletin: World Day to Fight Malaria. *Bol Epidemiol [Internet]*. p. 19–30. <https://www.gov.br/saude/pt-br/media/pdf/2021/fevereiro/22/boletim-epidemiologico-svs-17.pdf>. Accessed 20 Apr 2022.
- Brazilian Ministry of Health, Secretariat of Health Surveillance, Department of Immunization and Communicable Diseases, General Coordination of Surveillance of Zoonoses and Vector-Transmitted Diseases. Epidemiological Bulletin – Malaria: 2020 [Internet]. Nov. 2020. p. 118. [https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/boletins-epidemiologicos/especiais/2020/boletim\\_especial\\_malaria\\_1dez20\\_final.pdf](https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/boletins-epidemiologicos/especiais/2020/boletim_especial_malaria_1dez20_final.pdf). Accessed 20 Apr 2022.
- Brazilian Ministry of Health, Secretariat of Health Surveillance, General Coordination for the Development of Epidemiology in Services. Health Surveillance Guide. Single Vol. Ed. – Brasília: Brazilian Ministry of Health, 2019. p. 740.
- Malaria treatment guide in Brazil. Brazilian Ministry of Health, Secretariat of Health Surveillance, Department of Immunization and Communicable Diseases. Brasília: Brazilian Ministry of Health, 2020.
- Brazilian Institute of Geography and Statistics. Legal Amazon: What it is [online]. 2021. <https://www.ibge.gov.br/geociencias/cartas-e-mapas/mapas-regionais/15819-amazonia-legal.html?=&t=acesso-ao-produto>. Accessed 20 Apr 2022.
- R Core Team (2021) A: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. Accessed 20 Apr 2022.
- Swapnarekha H, Behera HS, Nayak J, Naik B, Kumar PS. Multiplicative Holt-Winter model for trend analysis and forecasting of COVID-19 spread in India. *SN Comput Sci*. 2021;2:416.
- Koehler AB, Snyder RD, Ord JK. Forecasting models and prediction intervals for the multiplicative Holt-Winters method. *Int J Forecast*. 2001;17:269–86.
- Machado RLD, Couto AAD, Cavasini CE, Calvosa VSP. Malária em região extra-Amazônica: situação no Estado de Santa Catarina. *Rev Soc Bras Med Trop*. 2003;36:581–6.
- Carlos BC, Rona LDP, Christophides GK, Souza-Neto JA. A comprehensive analysis of malaria transmission in Brazil. *Pathog Glob Health*. 2019;113:1–13.
- Tableau P. Data for Citizen by Malaria—Brazil [ONLINE]. [Internet]. [cited 2021 Jun 9]. <https://publictableau.com/app/profile/malaria.brasil>. Accessed 9 Jun 2021.
- Pina-Costa A, Brasil P, Di Santi SM, de Araujo MP, Suárez-Mutis MC, Santelli AC, et al. Malaria in Brazil: what happens outside the Amazonian endemic region. *Mem Inst Oswaldo Cruz*. 2014;109:618–33.
- Ahmed S, Reithinger R, Kaptoge SK, Ngondi JM. Travel is a key risk factor for malaria transmission in pre-elimination settings in sub-Saharan Africa: a review of the literature and meta-analysis. *Am J Trop Med Hyg*. 2020;103:1380–7.
- Massad E, Laporta GZ, Conn JE, Chaves LS, Bergo ES, Figueira EAG, et al. The risk of malaria infection for travelers visiting the Brazilian Amazonian region: a mathematical modeling approach. *Travel Med Infect Dis*. 2020;37:101792.

24. Massad E, Behrens RH, Burattini MN, Coutinho FA. Modeling the risk of malaria for travelers to areas with stable malaria transmission. *Malar J.* 2009;8:296.
25. Limongi JE, Costa DC, Carvalho LH, Lopes IS, Silva AA, Ferreira MS. Plasmodium ovale malaria in Brazil: report of an imported case with a prolonged incubation period. *J Infect Dev Ctries.* 2014;8:554–7.
26. Lüthi B, Schlegelhauf P. Risk factors associated with malaria deaths in travellers: a literature review. *Travel Med Infect Dis.* 2015;13:48–60.
27. Sarkar J, Murhekar MV, Shah NK, van Hutin Y. Risk factors for malaria deaths in Jalpaiguri district, West Bengal, India: evidence for further action. *Malar J.* 2009;8:133.
28. Lechthaler F, Matthys B, Lechthaler-Felber G, Likwela JL, Mavoko HM, Rika JM, et al. Trends in reported malaria cases and the effects of malaria control in the Democratic Republic of the Congo. *PLoS ONE.* 2019;14:e0219853.
29. Lima MVM, Laporta GZ. Evaluation of prediction models for the occurrence of malaria in the state of Amapá, Brazil, 1997–2016: an ecological study. *Epidemiol Serv Saude.* 2021;30:e2020080 (in Portuguese).
30. Bright A, Glynn-Robinson AJ, Kane S, Wright R, Saul N. The effect of COVID-19 public health measures on nationally notifiable diseases in Australia: preliminary analysis. *Commun Dis Intell.* 2020;5:44.
31. Lai CC, Chen SY, Yen MY, Lee PI, Ko WC, Hsueh PR. The impact of the coronavirus disease 2019 epidemic on notifiable infectious diseases in Taiwan: a database analysis. *Travel Med Infect Dis.* 2021;40:101997.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year


At BMC, research is always in progress.

Learn more [blomedcentral.com/submissions](https://blomedcentral.com/submissions)



## 12. ANEXO V



### Artigo publicado – Aspectos metodológicos sobre *record linkage*

 RESS | REVISTA DO SUS

FERRAMENTAS PARA GESTÃO  
DA VIGILÂNCIA EM SAÚDE  
doi 10.1590/S2237-96222022000300004

## Procedimentos para vinculação de dados da saúde: aplicações na vigilância em saúde

*Procedures for health data linkage: applications in health surveillance*  
*Procedimientos para la vinculación de datos sanitarios: aplicaciones en la  
vigilancia en la salud*

Klauss Kleydmann Sabino Garcia<sup>1</sup> , Cristiano Barreto de Miranda<sup>2</sup>   
Flávia Nogueira e Ferreira de Sousa<sup>1</sup> 

<sup>1</sup>Universidade de Brasília, Núcleo de Medicina Tropical, Brasília, DF, Brasil  
<sup>2</sup>Universidade de São Paulo, Faculdade de Saúde Pública, São Paulo, SP, Brasil


---

### RESUMO

**Objetivo:** Apresentar metodologia padronizada para vinculação de diferentes bancos de dados em saúde pública. **Métodos:** Artigo de revisão metodológica, com descrição específica de processos de tratamento de dados para vinculação (*linkage*) determinística entre bancos de dados estruturados. Instruiu-se como tratar os dados, selecionar chaves de vinculação e vincular os bancos, utilizando-se dois bancos de dados simulados no *software R*. **Resultados:** Foram apresentados os comandos utilizados para a vinculação determinística, do tipo *inner\_join*. O processo de vinculação resultou em um banco de dados com 40.108 pares ao se utilizar apenas a chave "Nome". Com a adição da segunda chave, "Nome da mãe", o resultado caiu para 112 pares. Ao adicionar a terceira chave, "Data de nascimento", apenas dois pares foram identificados. **Conclusão:** A vinculação de bancos de dados e suas análises são ferramentas válidas e úteis para os serviços de saúde, no apoio a ações de vigilância em saúde.

**Palavras-chave:** Análise de dados; Epidemiologia; Saúde Coletiva; Vigilância em Saúde Pública.

---

 Epidemiologia e Serviços de Saúde, Brasília, 31(3):e20211272, 2022 

## INTRODUÇÃO

A vigilância em saúde é definida como um processo contínuo e sistemático de coleta, consolidação, análise de dados e disseminação de informações sobre eventos relacionados à saúde, com vistas ao planejamento e implementação de medidas de saúde pública que incluem a regulação, intervenção e atuação em condicionantes e determinantes da saúde, para proteção e promoção da saúde da população, prevenção e controle de riscos, agravos e doenças.<sup>1</sup>

A integração dos sistemas de informações em saúde (SIS) é uma importante estratégia descrita na Política Nacional de Vigilância em Saúde (PNVS), no sentido de contribuir com o aprimoramento e consolidação da gestão da vigilância em saúde, especialmente nas atividades de planejamento, monitoramento e avaliação das ações de vigilância, em tempo oportuno.<sup>1</sup> Os SIS são ferramentas tecnológicas que auxiliam o processamento de dados coletados em serviços de saúde e outros locais, gerando informações úteis para a compreensão dos problemas e subsídio à tomada de decisão no âmbito das políticas e do cuidado em saúde.<sup>2,3</sup>

No Brasil, o Ministério da Saúde é responsável pela gestão dos SIS de base nacional. Entre os diversos SIS, os mais utilizados são: o Sistema de Informações sobre Mortalidade (SIM), o Sistema de Informações sobre Nascidos Vivos (Sinasc), o Sistema de Informação de Agravos de Notificação (Sinan), o Sistema de Informações Hospitalares do Sistema Único de Saúde (SIH/SUS), o Sistema de Informação de Vigilância Epidemiológica (Sivep) e, recentemente, o sistema e-SUS Notifica. Apesar da disponibilidade e aperfeiçoamento de diversos SIS nas últimas décadas, a interoperabilidade entre eles ainda não ocorre nos setores da vigilância em saúde.<sup>2</sup>

Frente à necessidade de se trabalhar com informações qualificadas, que mantenham integração entre os diferentes SIS, o relacionamento de base de dados apresenta-se como uma estratégia viável, com o propósito de

Contribuições do estudo	
<b>Principais resultados</b>	O principal resultado deste trabalho é o modelo metodológico padronizado de vinculação de dados de diferentes sistemas de informações em saúde pública.
<b>Implicações para os serviços</b>	Este trabalho permitirá a replicação de métodos padronizados de vinculação de dados, com embasamento técnico-científico, para as análises de situação em saúde.
<b>Perspectivas</b>	Espera-se que os procedimentos apresentados sejam utilizados na íntegra ou como modelos para processos de vinculação de dados nos serviços de saúde e instituições de pesquisa ou ensino, para aprimoramento das análises epidemiológicas.

materializar essa conexão de diferentes informações.<sup>4</sup> A técnica de *linkage*, de relativo baixo custo operacional, permite a recuperação de registros incompletos ou inconsistentes e, com isso, melhora a completude e a confiabilidade das informações disponibilizadas pelos SIS.<sup>5</sup>

Embora estudos anteriores evidenciem as vantagens do emprego de *linkage* para a análise da qualidade de diversos SIS,<sup>6-9</sup> há uma escassez de estudos metodológicos que apresentem os procedimentos necessários à realização dessa técnica. Assim, este manuscrito justifica-se pela importância de disseminar metodologias utilizadas na vigilância em saúde, na padronização metodológica de vinculação de dados e no fornecimento de modelos de análise possíveis de serem adaptados a diferentes realidades e áreas de conhecimento da saúde.

O objetivo deste trabalho foi apresentar um método padronizado para a vinculação de diferentes bancos de dados em saúde pública.

## MÉTODOS

Artigo de revisão metodológica, sobre vinculação de bancos de dados de diferentes sistemas de informações, considerando procedimentos técnicos realizados no âmbito da vigilância em saúde do Ministério da Saúde. Este material foi desenvolvido para utilização no *software* R, a partir de bancos de dados simulados com informações hipotéticas sobre nome do paciente, nome da mãe e sua data de nascimento.

Os métodos descritos a seguir são replicáveis para quaisquer dados de diferentes SIS, e estes são focados no processo de vinculação determinística, que realiza a classificação dos possíveis pares com base em regras de combinação que prezam pelo pareamento de registros idênticos.<sup>10</sup>

### Preparação dos bancos de dados

É importante que, antes de iniciar o processo de vinculação de dados, sejam definidas as variáveis a serem utilizadas para a análise dos dados vinculados. Essa definição subsidiará o processo de tratamento dos dados, pois, ao se trabalhar com bancos de dados de grande magnitude, pode ser necessário diminuir o tamanho do banco para que a máquina (computador) processe os dados com maior facilidade, evitando-se problemas de *performance*. Além disso, o *software* R pode apresentar mau funcionamento, por limitação de memória interna do programa ao se utilizar em bancos de dados de grande magnitude. Assim, é recomendável trabalhar com objetividade, utilizando-se apenas as variáveis de interesse, para melhor eficiência do processo de vinculação.

### Software R para vinculação

Sugere-se a utilização do *software* R no processo de desenvolvimento do roteiro de tratamento e vinculação de dados.

Será necessário realizar a instalação dos pacotes *lubridate*<sup>11</sup> (que permite o tratamento

de variáveis do tipo "data", como, por exemplo, "Data de notificação"), *abjutils*<sup>12</sup> (permite a remoção de acentos gráficos) e *tidyverse*;<sup>13</sup> para simular dados nominais, sugere-se o pacote e função "*randomNames*".<sup>14</sup>

O pacote *tidyverse*<sup>13</sup> engloba outros pacotes, mais frequentemente utilizados no tratamento e análise de dados, como: *ggplot2* (permite a produção de gráficos), *dplyr* (manipulação de dados), *tidyr* (organização dos dados), *readr* (leitura de arquivos .csv), *purrr* (permite trabalhar com funções e vetores), *tibble* (leitura de bancos de dados em formato de texto .txt), *stringr* (permite trabalhar com variáveis nominais) e *forcats* (tratamento de variáveis categóricas) – comando: `install.packages("tidyverse")`. Caso o *software* seja fechado, deve-se utilizar a função *library* para carregamento dos pacotes de interesse, já instalados – comando: `library(tidyverse)`.

### Tratamento de dados

É importante que as etapas descritas a seguir sejam replicadas seguindo a mesma ordem de instruções deste passo a passo, para evitar erros.

Inicialmente, deve-se definir o diretório de trabalho (pasta de arquivos) com a função *setwd*. Nesse diretório, devem ser inseridos os bancos de dados que serão vinculados – comando: `setwd("C:/Users/User/Desktop/Pasta de trabalho")`.

Para importação dos bancos de dados ao ambiente do R, sugere-se a utilização da função *read.csv2* – que importa arquivos do tipo csv com separador de ";" ao invés de ",". Para bancos de dados no formato de texto (txt), deve-se utilizar a função *read.table* e definir a primeira linha como um cabeçalho de variáveis, o separador de células para ";" e o marcador decimal para ".". Caso sejam utilizados arquivos do tipo xls ou xlsx, deve-se utilizar o comando `readxl::read_excel("arquivo.xlsx")`. Para arquivos do tipo dbf ou dbc,<sup>15</sup> comuns para dados do Sinan, é necessária a instalação dos pacotes "foreign"<sup>16</sup> e "read.dbc",<sup>17</sup> e utilização

dos comandos `read.dbf("arquivo.dbf")` e `read.dbc::read.dbc("arquivo.dbc")` respectivamente.

Sugere-se a atribuição do primeiro banco de dados importado a um objeto de armazenamento do R, denominado "Banco\_1", e o segundo banco a outro objeto, denominado "Banco\_2" – comando: `Banco_1 <- read.csv2("Banco de dados 1.csv")`; ou `read.table("Banco_1.txt", header = TRUE, sep = ";", dec = ",")`.

Para cada banco de dados, sugere-se a exclusão das variáveis que não serão utilizadas, por meio do comando `"Banco_1$variavel <- NULL"`. Quando a quantidade de variáveis a serem retiradas for grande, sugere-se, ao invés da exclusão, a seleção das variáveis de interesse, por meio da função `select` e atribuição desse novo conjunto ao mesmo objeto anteriormente criado – exemplo: `Banco_1 <- select(Banco_1, variavel1, variavel3, variavel5)`.

É essencial que sejam mantidas as variáveis que serão utilizadas como chaves para a vinculação, junto às que serão analisadas. Sugere-se a utilização das variáveis referentes ao nome da pessoa notificada, CPF e data de nascimento, e o nome da mãe, por serem as chaves mais frequentemente utilizadas na literatura científica. Apesar de incomum em bancos de dados dos SIS, a utilização da variável "CPF" pode ser entendida como uma variável que possui pontos e hífen em sua estrutura. Esta situação configura o CPF como uma variável do tipo *character*, e, para sua utilização como uma variável numérica (*numeric* ou *integer*), é necessário limpar esses pontos e hífens das observações. Além disso, o preenchimento do CPF pode ser feito de diferentes formas, nas notificações desses SIS, contendo ou não hífens e pontos. Devido a essa falta de padronização no preenchimento de variáveis desse tipo, sugere-se a remoção desses sinais para o melhor aproveitamento dessas chaves.<sup>19-20</sup>

Para a padronização das variáveis "nome" e "CPF", sugere-se que estas sejam do tipo *character* (texto). Quando for necessária a conversão dessas variáveis, deve-se utilizar o comando `as.character(Banco_1$variavel1)`. Para

o CPF, deve-se retirar os pontos "." e traços "-" (comando: `Banco_1$CPF <- str_replace_all(Banco_1$CPF, "\\.|-", "")`), e adicionar dígitos zero "0" à esquerda do texto até completar 11 dígitos (comando: `Banco_1$CPF <- str_pad(RAIS Banco_1$CPF, 11, pad="0")`).

Para a variável "data do nascimento", é preciso conferir se seu formato aparece como "Date" (data) no ambiente R. A checagem do tipo da variável pode ser feita com a função `class` – comando: `class(Banco_1$Data_nascimento)`.

Datas, no R, são exibidas da seguinte forma: 2020-12-31 (ano-mês-dia). Caso a variável não esteja no formato "Date", cumpre convertê-la para este formato por meio de funções do pacote *lubridate*,<sup>21</sup> por exemplo, via comando: `Banco_1$Data_nascimento <- as.Date(Banco_1$Data_nascimento)`. Outras variáveis deste comando podem ser utilizadas, como `dmy` (para datas escritas em dia-mês-ano) ou `ymd` (para datas em ano-mês-dia).

Para as variáveis de "Nome", deve-se, nesta ordem, substituir as letras acentuadas por letras sem acentuação gráfica – por exemplo; substituir "Á" por "A", "Ô" por "O" e demais opções. Podem-se também retirar conectores de sobrenomes, como "E", "A", "DA", "DE", "DO", "DAS", "DES" e "DOS" – por exemplo; "MARIA DOS ANJOS" torna-se "MARIA ANJOS". Sugere-se a transformação das letras dos nomes para maiúsculas, pois o *software* faz distinção entre letras maiúsculas e minúsculas. Para isto, pode-se utilizar o comando `toupper(Banco_1$Nome)`. Para a substituição de letras, utiliza-se a função `str_replace_all` – comando: `Banco_1$Nome <- str_replace_all(Banco_1$Nome, " DAS | DE | DO | DA |,|:|-| DE | E | DO | DAS | DOS | ", " ")`.

Após esta substituição, sugere-se a exclusão dos espaços em branco entre os nomes, pois é possível que se encontrem espaços duplos entre os nomes, e isto impediria processos de vinculação determinística – comando: `Banco_1$Nome <- str_replace_all(Banco_1$Nome, " ", "")`.

Esse processo de padronização das variáveis de nome do notificado, CPF, data de nascimento e nome da mãe considera a possibilidade de pequenas inconsistências no preenchimento dessas informações pelos notificantes, nos diferentes sistemas de informações.

Após a limpeza e padronização dessas informações, é possível explorar o banco de dados para verificação de registros duplicados. A função *distinct* pode ser utilizada para excluir registros duplicados; porém, a exclusão ou não desses registros duplicados deve ser discutida antecipadamente, considerando-se o objetivo do resultado da vinculação de dados.

Previamente à vinculação, também deve ser estabelecido quantas chaves serão utilizadas no processo. A combinação de diferentes chaves, como a chave "Nome\_paciente" combinada com a chave "Data\_nascimento", acarretará um processo de vinculação mais específico, prezando-se por maior número de registros verdadeiros-positivos. A utilização de apenas uma chave, como "Nome\_paciente", apresentará resultados mais sensíveis, com maior presença de pareamentos falsos-positivos. Destaca-se ainda que o *software* R realiza análises combinatórias para os possíveis pares; logo, se há cinco pessoas com o nome

"MARIA DOS ANJOS" em cada banco de dados, o *software* retornará um banco vinculado com 25 resultados. Devido a isso, sugere-se sempre a utilização de pelo menos duas variáveis-chave de vinculação.

A decisão de se utilizar mais de duas chaves para a vinculação deve estar embasada em análises de qualidade dos bancos de dados, pois o processo determinístico pode ser afetado por dados digitados incorretamente; por exemplo, quando se encontram datas de nascimento diferentes, mas, mediante análises manuais, o pesquisador consegue identificar serem as mesmas, embora tenham sido inseridas com algum erro de digitação. Dessa forma, pode ser interessante utilizar métodos mais sensíveis, desde que acompanhados por limpezas manuais de possíveis pareamentos verdadeiros-negativos (Quadro 1).

#### Vinculação de dados

A vinculação determinística identifica pares de registros concordantes, a partir de um determinado conjunto de regras.<sup>21</sup> Ela é indicada quando os bancos de dados a serem trabalhados apresentarem uma variável identificadora comum ou um conjunto de variáveis com boa qualidade de preenchimento.<sup>10,22,23</sup>

**Quadro 1 – Exemplos de variáveis-chave utilizadas para vinculação, e possíveis resultados**

Banco de dados 1			Banco de dados 2			Resultado	Avaliação
Nome_paciente <sup>a</sup>	Data_nascimento <sup>a</sup>	Nome_mae	Nome_paciente <sup>a</sup>	Data_nascimento <sup>a</sup>	Nome_mae		
MariaAnjos	27-09-1994	LurdesSilva	MariaAnjos	27-09-1994	LurdesSilva	Pareado	Verdadeiro positivo
MariaAnjos	27-09-1994	LurdesSilva	MariaAnjos	27-09-1994	AnaCleide	Pareado	Falso positivo
MariaAnjos	27-09-1994	LurdesSilva	MariaAnjos	13-03-2001	MariaDores	Não pareado	Verdadeiro negativo
MariaAnjos	27-09-1994	LurdesSilva	MariaAnjos	27-09-1894	LurdesSilva	Não pareado	Falso negativo

a) Campo utilizado como chave.

Utilizando-se o *software* R, é possível realizar diferentes formas de vinculação. As mais frequentes, descritas a seguir, são exemplificadas no Quadro 2. As funções a serem utilizadas são "left\_join", "inner\_join", "full\_join" e "semi\_join".<sup>24</sup> A função *left\_join* alimenta o Banco de dados 1 com informações do Banco de dados 2. A função *inner\_join* retorna apenas as informações com chaves comuns aos dois bancos, ou seja, apenas os registros concordantes. A função *full\_join* junta ambos os bancos de dados, sem perda de informações. A função *anti\_join* retorna os registros do Banco de dados 1 que não encontraram pareamento com o Banco de dados 2 (Figura 1).

Para que o funcionamento das funções descritas previamente seja facilitado, indica-se que as variáveis-chave para a vinculação recebam exatamente o mesmo nome; caso contrário, as funções precisarão da especificação dos diferentes nomes das chaves.

Após a conclusão da vinculação, o banco de dados estará pronto para ser explorado e analisado. Os principais pacotes para análise desses dados já terão sido carregados indiretamente, pelo carregamento do pacote *tidyverse*.<sup>13</sup>

#### Simulação do processo de vinculação determinística

Para a simulação da vinculação determinística, foram criados dois bancos de dados (*data frames*) por meio da função "randomNames" do pacote "randomNames".<sup>14</sup> O Banco\_1 foi criado com 10 mil observações, e o Banco\_2, com 500 mil observações. Nestes bancos, foram incluídas variáveis simuladas para "Nome" (nome e sobrenome), "Nome da mãe" (apenas o primeiro nome), "Data de nascimento", "Nome do Banco" e outras duas variáveis teste (variável teste X, teste Y, Teste W, e Teste Z), totalizando seis variáveis em cada banco de dados. O *script* R para criação dos bancos simulados está disponível no material suplementar ([\[thub.com/kleydmann/Modelo\\\_Vinculacao\\\_deterministica.git\]\(https://github.com/kleydmann/Modelo\_Vinculacao\_deterministica.git\)\).](https://gi-</a></p>
</div>
<div data-bbox=)

## IMPLEMENTAÇÃO DA VINCULAÇÃO

Roteiro de dados:

Instalação e carregamento dos pacotes necessários:

```
install.packages("tidyverse")
install.packages("lubridate")
install.packages("abjutils")
library(tidyverse)
library(lubridate)
library(abjutils)
```

Definição de diretório:

```
setwd("C:/Users/User/Desktop/Pasta de trabalho")
```

Importação dos bancos de dados para o ambiente R:

```
Banco_1 <- read.csv2("Banco_1.csv")
Banco_2 <- read.csv2("Banco_2.csv")
```

Exclusão de variáveis, se necessário:

```
Banco_1$X_Y <- NULL
Banco_2$X_Z <- NULL
```

Seleção das variáveis de interesse, se necessário:

```
Banco_1 <- select( Banco_1, Nome, Nome_mae, DT_NASC, Banco, variavel_X)
Banco_2 <- select( Banco_2, Nome, Nome_mae, DT_NASC, Banco, variavel_Z)
```

Após o tratamento dos dados, o Banco\_1 e o Banco\_2 foram compostos por 10 mil e 500 mil observações, respectivamente, e cinco variáveis cada um.

Quadro 2 – Exemplos de funções para vinculação de dados do software R

Método	Quantidade de chaves	Funções
left_join	1	left_join(Banco_1, Banco_2, by = "Nome_paciente ")
	2	left_join(Banco_1, Banco_2, by = c("Nome_paciente ", "Data.de.Nascimento"))
	3	left_join(Banco_1, Banco_2, by = c("Nome_paciente ", "Data.de.Nascimento", "nome.da.mae"))
inner_join	1	inner_join(Banco_1, Banco_2, by = "Nome_paciente ")
	2	inner_join(Banco_1, Banco_2, by = c("Nome_paciente ", "Data.de.Nascimento"))
	3	inner_join(Banco_1, Banco_2, by = c("Nome_paciente ", "Data.de.Nascimento", "nome.da.mae"))
full_join	1	full_join(Banco_1, Banco_2, by = "Nome_paciente ")
	2	full_join(Banco_1, Banco_2, by = c("Nome_paciente ", "Data.de.Nascimento"))
	3	full_join(Banco_1, Banco_2, by = c("Nome_paciente ", "Data.de.Nascimento", "nome.da.mae"))
anti_join	1	anti_join(Banco_1, Banco_2, by = "Nome_paciente ")
	2	anti_join(Banco_1, Banco_2, by = c("Nome_paciente ", "Data.de.Nascimento"))
	3	anti_join(Banco_1, Banco_2, by = c("Nome_paciente ", "Data.de.Nascimento", "nome.da.mae"))

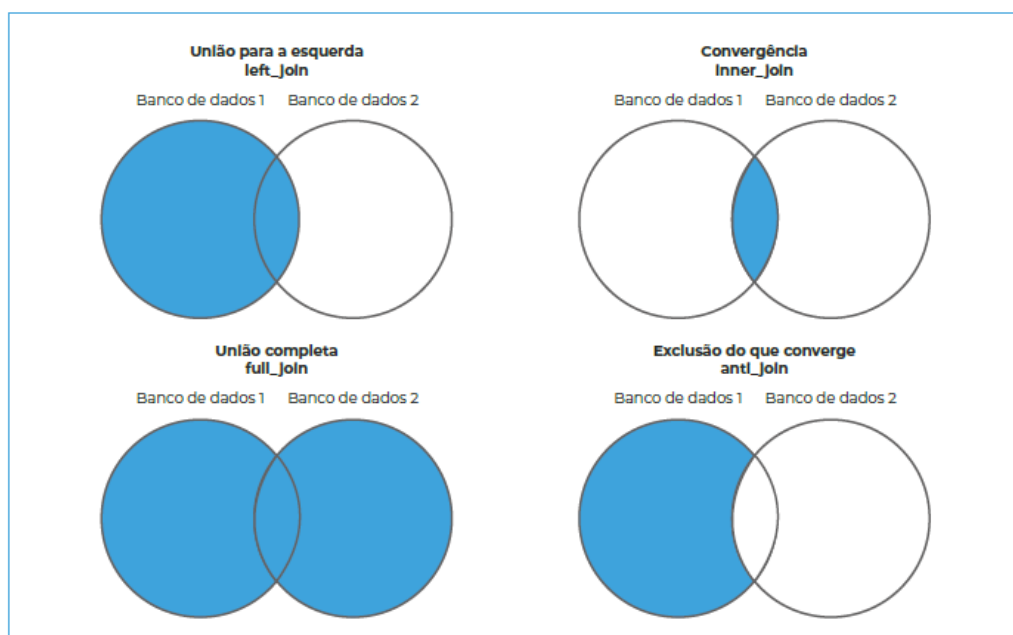


Figura 1 – Representação dos tipos de união de banco de dados, segundo funções do software R

Verificação da codificação das variáveis-chave:

```
> class(Banco_1$Nome)
[1] "character"
> class(Banco_1$Nome_mae)
[1] "character"
> class(Banco_1$DT_NASC)
[1] "Date"
> class(Banco_2$Nome)
[1] "character"
> class(Banco_2$Nome_mae)
[1] "character"
> class(Banco_2$DT_NASC)
[1] "Date"
```

Convertendo tipo da variável de "Data de nascimento":

```
Banco_1$DT_NASC <- ymd(Banco_1$DT_NASC)
Banco_2$DT_NASC <- ymd(Banco_2$DT_NASC)
```

Transformando as letras em maiúsculas:

```
Banco_1$Nome <- toupper(Banco_1$Nome)
Banco_1$Nome_mae <- toupper(Banco_1$Nome_mae)
Banco_2$Nome <- toupper(Banco_2$Nome)
Banco_2$Nome_mae <- toupper(Banco_2$Nome_mae)
```

Tratamento das variáveis "Nome" e "Nome\_mae" do Banco\_1 e Banco\_2 (remoção de palavras conectoras, preposições, acentos gráficos e espaços duplos):

```
Banco_1$Nome <- str_replace_all(Banco_1$Nome, "\\ | DA |,|:|-| DE | E | DO | DAS | DOS | ", "")
Banco_1$Nome <- rm_accent(Banco_1$Nome)
```

```
Banco_1$Nome_mae <- str_replace_all(Banco_1$Nome_mae, "\\ | DA |,|:|-| DE | E | DO | DAS | DOS | ", "")
Banco_1$Nome_mae <- rm_accent(Banco_1$Nome_mae)
Banco_2$Nome <- str_replace_all(Banco_2$Nome, "\\ | DA |,|:|-| DE | E | DO | DAS | DOS | ", "")
Banco_2$Nome <- rm_accent(Banco_2$Nome)
Banco_2$Nome_mae <- str_replace_all(Banco_2$Nome_mae, "\\ | DA |,|:|-| DE | E | DO | DAS | DOS | ", "")
Banco_2$Nome_mae <- rm_accent(Banco_2$Nome_mae)
```

Remoção de espaços em branco:

```
Banco_1$Nome <- str_replace_all(Banco_1$Nome, " ", "")
Banco_1$Nome_mae <- str_replace_all(Banco_1$Nome_mae, " ", "")
Banco_2$Nome <- str_replace_all(Banco_2$Nome, " ", "")
Banco_2$Nome_mae <- str_replace_all(Banco_2$Nome_mae, " ", "")
```

Vinculação determinística por método *inner\_join* (apenas registros congruentes em ambas as bases)

Utilização de 1 chave:

```
Banco_vinculado_1_chave <- inner_join(Banco_1, Banco_2, by = "Nome")
```

Utilização de 2 chaves:

```
Banco_vinculado_2_chaves <- inner_join(Banco_1, Banco_2, by = c("Nome", "Nome_mae"))
```

Utilização de 3 chaves:

```
Banco_vinculado_3_chaves <- inner_
```

```
join(Banco_1,Banco_2, by = c("Nome","Nome_
mae","DT_NASC"))
```

O processo de vinculação determinística *inner\_join* utilizando uma única variável-chave ("Nome") resultou em um banco de dados com 40.108 possíveis pares e nove variáveis. A utilização de apenas uma chave resultou em maior quantidade de possíveis pares do que o total de registros existentes no Banco de dados 1, isto devido às análises combinatórias que o *software* R realiza durante a vinculação.

Ao se utilizarem duas variáveis-chave ("Nome" e "Nome da mãe"), foram encontrados 112 pares (1,12%). Ao se adicionar a terceira variável-chave, "Data de nascimento", apenas dois (0,02%) pares foram identificados.

## DISCUSSÃO

Embora o relacionamento de banco de dados, também conhecido como *linkage*, tenha sido descrito em diversos estudos<sup>25-28</sup> como uma técnica relevante para melhorar a qualidade das informações em saúde, sua utilização ainda não é amplamente difundida nos ambientes da vigilância em saúde.

Entre as principais vantagens do *linkage* para os serviços de saúde, estão (i) a possibilidade de aprimorar as informações provenientes dos sistemas de informações, com resgate de informações incompletas, e (ii) a identificação de erros no preenchimento dos dados. Maia, Souza e Mendes<sup>25</sup> verificaram que a técnica de *linkage* contribuiu para o aprimoramento da qualidade dos dados de mortalidade infantil em cinco cidades brasileiras, com 92% dos campos incompletos do SIM e do Sinasc recuperados. Dados sobre tuberculose no município do Rio de Janeiro foram qualificados, com redução de inconsistências no banco de dados após *linkage* com utilização dos bancos de dados do SIM e dados sobre aids (Sinan).<sup>28</sup>

Quanto à abordagem metodológica deste trabalho, não se utilizou a abordagem proba-

bilística, por esta requerer definições de regras de pareamento e delimitação de pontos de corte para os índices de similaridade, fazendo-se assim necessário o desenvolvimento de um estudo específico para tais técnicas.<sup>23</sup>

As estratégias de *linkage* abordadas neste estudo tendem a apresentar um baixo número de identificação de pares verdadeiros, ao se utilizarem muitas variáveis-chave no processo de pareamento. Sendo assim, talvez seja interessante utilizar códigos fonéticos combinados com outras variáveis, uma alternativa possível para aumentar a sensibilidade do método.

No entanto, é essencial a definição de conjuntos de regras que permitam a utilização de chaves de vinculação que sejam precisas, estáveis ao longo do tempo e comuns às diferentes bases de dados de interesse.<sup>20</sup>

As quatro diferentes funções utilizadas para o *linkage* neste estudo permitem diferentes formas de exploração dos dados após a vinculação. O método *left\_join* e o método *inner\_join* são mais frequentemente utilizados, pois adotam o primeiro banco de dados como referência e o segundo como fonte de novas informações. Assim, o método *left\_join* pode, por exemplo, ser utilizado para a complementação de informações referentes a "Ocupação", uma variável que não se encontra em todas as fichas de notificação do Sinan, embora esteja presente em alguns sistemas de informações do Ministério do Trabalho e Previdência. O método *inner\_join*, por exemplo, pode ser utilizado para parear informações de notificações de casos do Sinan com notificações de óbitos presentes no SIM, permitindo a análise específica de casos que evoluíram a óbito.

A limpeza prévia dos dados a serem pareados permite ampliar a acurácia geral do algoritmo de *linkage* empregado, dado que alguns bancos apresentam má qualidade de dados.

É importante mencionar que os procedimentos descritos neste trabalho possuem um nível de operacionalização e replicabilidade de maior simplicidade. Porém, a depender da

complexidade da vinculação que se planeja, outros pacotes, como o "RecordLinkage",<sup>29</sup> e métodos podem ser mais adequados.

De todo modo, é preciso estar atento aos pareamentos entre bancos que possuam registros duplicados ou registros diferentes para uma mesma pessoa, pois as funções descritas neste trabalho resultam em diferentes possibilidades de análises combinatórias frente a registros idênticos.

As limitações descritas aqui, sobre o método e as técnicas gerais de vinculação de dados, estão principalmente relacionadas à possibilidade de erros sistemáticos provenientes da utilização de bancos de dados secundários, desde que a utilização de bancos de dados de grande magnitude torna onerosa a tarefa de conferência manual de possíveis pares falso-positivos. Ademais, a limitação do *software* R no gerenciamento de memória interna é um aspecto que pode dificultar a execução da vinculação. Logo, é essencial que se discuta previamente o processo de vinculação, visando diminuir a possibilidade de vieses de seleção e de problemas operacionais do *software*.

A qualidade do banco de dados vinculado é dependente da qualidade dos bancos de dados originais. Dessa forma, é importante que seja feita uma avaliação prévia da qualidade dos dados, para se conhecer as deficiências e limitações de cada banco de dados e, se possível,

corrigi-las antes da vinculação. Esta é uma etapa importante no pré-processamento analítico, em análise de dados.

## CONSIDERAÇÕES FINAIS

As análises de bancos de dados vinculados apresentam-se como uma ferramenta válida e útil aos serviços de saúde, principalmente os que configuram as esferas estaduais e federais de governo.

Estes métodos permitem a realização de estudos descritivos e guardam potencial para subsidiar estudos analíticos, bem como fornecer informações que podem contribuir no desenvolvimento de ações estratégicas em saúde e no fomento de políticas públicas de saúde voltadas a populações mais vulneráveis. A incorporação do método *linkage* na rotina dos serviços de saúde pode ser uma ferramenta a se utilizar, contribuindo para a implantação de ações mais adequadas, visando à melhoria da vigilância em saúde.

Por fim, a necessidade de se implantarem processos de vinculação de dados destaca a fragilidade da interoperabilidade entre os sistemas de informações governamentais. Tal deficiência precisa ser trabalhada, promovendo melhor integração entre sistemas de informações, tanto da saúde como de outras áreas do Estado.

### CONTRIBUIÇÃO DOS AUTORES

Garcia KKS contribuiu na concepção e delineamento do estudo, análise e interpretação dos resultados, redação e revisão crítica do conteúdo do manuscrito. Garcia KKS, Miranda CB e Sousa FNS contribuíram na concepção e delineamento do estudo, e revisão crítica do conteúdo do manuscrito. Todos os autores aprovaram a versão final do manuscrito e são responsáveis por todos os seus aspectos, incluindo a garantia de sua precisão e integridade.

### CONFLITOS DE INTERESSE

Os autores declararam não haver conflitos de interesse.

**Correspondência:** Klauss Kleydmann Sabino Garcia | E-mail: kleydmann25@gmail.com

**Recebido em:** 16/02/2022 | **Aprovado em:** 08/07/2022

**Editora associada:** Elisângela Aparecida da Silva Lizzi 

## REFERÊNCIAS

1. Ministério da Saúde (BR). Conselho Nacional de Saúde. Resolução n. 588, de 12 de julho de 2018. Instituição da Política Nacional de Vigilância em Saúde. 2018. Diário Oficial da União, Brasília (DF), 2018 ago 13 [citado em 02 de maio de 2022]. Seção 1:87. Disponível em: [https://www.in.gov.br/materia/-/asset\\_publisher/Kujrw0TZC2Mb/content/id/36469447/doi-10.1136/bmjopen-2020-037885](https://www.in.gov.br/materia/-/asset_publisher/Kujrw0TZC2Mb/content/id/36469447/doi-10.1136/bmjopen-2020-037885)
2. Coelho Neto GC, Chioro A. Afinal, quantos Sistemas de Informação em Saúde de base nacional existem no Brasil?. *Cad Saude Publica*. 2021;37(7):e00182119. doi: 10.1590/0102-311X00182119
3. World Health Organization. *Monitoring the building blocks of health systems: a handbook of indicators and their measurement strategies*. Geneva: World Health Organization; 2010. 92 p.
4. Coathup V, Macfarlane A, Quigley M. Linkage of maternity hospital episode statistics birth records to birth registration and notification records for births in England 2005–2006: quality assurance of linkage. *BMJ Open* 2020;10(10): e037885. doi: 10.1136/bmjopen-2020-037885
5. Paes NA, Santos CSA, Coutinho TDF. Qualidade dos registros de óbitos infantis para espaços regionalizados: um percurso metodológico. *Rev Bras Epidemiol*. 2021;24:e210016. doi: 10.1590/1980-549720210016
6. Lima SVMA, Cruz LZ, Araújo DC, Santos AD, Queiroz AAFLN, Araújo KCGM, et al. Quality of tuberculosis information systems after record linkage. *Rev Bras Enferm*. 2020;73(suppl 5):e20200536. doi: 10.1590/0034-7167-2020-0536
7. Szwarcwald CL, Leal MC, Esteves-Pereira AP, Almeida WS, Frias PG, Damascena GN, et al. Avaliação das informações do Sistema de Informações sobre Nascidos Vivos (SINASC), Brasil. *Cad Saude Publica*. 2019;35(10):e00214918. doi: 10.1590/0102-311X00214918
8. Almeida ABM, Silva ZP. Uso de linkage para análise de completude e concordância de óbitos por sífilis congênita na Região Metropolitana de São Paulo, 2010-2017: estudo descritivo. *Epidemiol Serv Saude*. 2021;30(4):e2021167. doi: 10.1590/S1679-49742021000400013
9. Marques LJP, Pimentel DR, Oliveira CM, Vilela MBR, Frias PG, Bonfim CV. Concordância da causa básica dos óbitos infantis antes e após a investigação no Recife, Pernambuco, 2014. *Epidemiol Serv Saude*. 2018;27(1):e20170557. doi: 10.5123/S1679-49742018000100007
10. Machado CJ. A literature review of record linkage procedures focusing on infant health outcomes. *Cad Saude Publica*. 2004;20(2):362-71. doi: 10.1590/S0102-311X2004000200003
11. Spinu V, Grolemond G, Wickham H. *Make Dealing with Dates a Little Easier* [Internet]. New York: Datacamp; 2021 [update 2021 Oct 7; cited 2022 May 2]. Available from: <https://cran.r-project.org/web/packages/lubridate/lubridate.pdf>
12. Lente C, Trecenti J, Witkoski K, Associação Brasileira de Jurimetria. *Useful Tools for Jurimetrical Analysis Used by the Brazilian Jurimetrics Association* [Internet]. New York: Datacamp; 2022 [update 2022 Feb 1; cited 2022 May 2]. Available from: <https://cran.rstudio.com/web/packages/abjutils/abjutils.pdf>
13. Wickham H. *Easily Install and Load the 'Tidyverse'* [Internet]. New York: Datacamp; 2021 [update 2021 Apr 15; cited 2022 May 2]. Available from: <https://cran.r-project.org/web/packages/tidyverse/tidyverse.pdf>

14. Betebenner DW. Generate Random Given and Surnames [Internet]. New York: Datacamp; 2022 [update 2022 Apr 22; cited 2022 May 12]. Available from: <https://cran.r-project.org/web/packages/randomNames/randomNames.pdf>
15. Saldanha RF, Bastos RR, Barcellos C. Microdatasus: pacote para download e pré-processamento de microdados do Departamento de Informática do SUS (DATASUS). *Cad Saude Publica*. 2019;35(9):e00032419. doi: 10.1590/0102-311X00032419
16. R Core Team, Bivand R, Carey VJ, DebRoy S, Eglen S, Guha R, et al. Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', ... [Internet]. Datacamp. New York: Datacamp; 2022 [update 2022 Jan 13; cited 2022 May 2]. Available from: <https://cran.r-project.org/web/packages/foreign/foreign.pdf>
17. Petruzalek D. Read Data Stored in DBC (Compressed DBF) Files [Internet]. New York: Datacamp; 2022 [update 2022 May 2; cited 2022 May 2]. Available from: <https://cran.r-project.org/web/packages/read.dbc/read.dbc.pdf>
18. Gill L. Methodology and applications of building a national file of health and mortality data. *Austrian Journal of Statistics*. 2004;33(1&2):101-24. doi: 10.17713/ajs.v33i1&2.433
19. Churches T, Christen P, Lim K, Zhu JX. Preparation of name and address data for record linkage using hidden Markov models. *BMC Med Inform Decis Mak*. 2002;2:9. doi:10.1186/1472-6947-2-9
20. Yaohao P, Mation LF. O desafio do pareamento de grandes bases de dados: mapeamento de métodos de record linkage probabilístico e diagnóstico de sua viabilidade empírica. Instituto de Pesquisa Econômica Aplicada. Brasília: Rio de Janeiro: Instituto de Pesquisa Econômica Aplicada; 2018. 48 p.
21. Brustulin R, Marson PG. Inclusão de etapa de pós-processamento determinístico para o aumento de performance do relacionamento (linkage) probabilístico. *Cad Saude Publica*. 2018;34(6):e00088117. doi: 10.1590/0102-311X00088117
22. Queiroz OV, Júnior AAG, Machado CJ, Andrade EIG, Júnior WM, Acurcio FA, et al. Relacionamento de registros de grandes bases de dados: estimativa de parâmetros e validação dos resultados, aplicados ao relacionamento dos registros das autorizações de procedimentos ambulatoriais de alta complexidade com os registros de sistema de informações hospitalares. *Cad Saúde Colet*. 2010;18(2):298-308.
23. Peres SV, Latorre MRDO, Tanaka LF, Michels FAS, Teixeira MLP, Coeli CM, et al. Melhora na qualidade e completude da base de dados do Registro de Câncer de Base Populacional do município de São Paulo: uso das técnicas de linkage. *Rev Bras Epidemiol*. 2016;19(4):753-65. doi: 10.1590/1980-5497201600040006
24. Devmedia. SQL JOIN: entenda como funciona o retorno de dados [Internet]. [S.l.]: Plataforma Devmedia; 2014 [citado 2020 Fev 11]. Disponível em <https://www.devmedia.com.br/sql-join-entenda-como-funciona-o-retorno-dos-dados/31006>
25. Maia LTS, Souza WV, Mendes ACG. A contribuição do linkage entre o SIM e SINASC para a melhoria das informações da mortalidade infantil em cinco cidades brasileiras. *Rev Bras Saude Mater Infant*. 2015;15(1):57-66. doi: 10.1590/S1519-38292015000100005
26. Bronhara B, Conde WL, Liciardi DC, França-Junior I. Vinculação determinística de Bancos de Dados sobre mortalidade por Aids. *Rev Bras Epidemiol*. 2008;11(4):709-13. doi: 10.1590/S1415-790X2008000400017
27. Rabelo ACL, Amâncio FF, Oiko CSF, Ferraz ML, Carneiro M. Caracterização dos casos confirmados de dengue por meio da técnica de linkage de bancos de dados, para avaliar a circulação viral em Belo Horizonte, 2009-2014. *Epidemiol Serv Saude*. 2020;29(3):e2019354. doi: 10.5123/S1679-49742020000300016

28. Rocha MS, Oliveira GP, Guillen LCT, Coeli CM, Saraceni V, Pinheiro RS. Uso de linkage entre diferentes bases de dados para qualificação de variáveis do Sinan-TB e a partir de regras de scripting. *Cad Saude Publica*. 2019;35(12):e00074318. doi: 10.1590/0102-311X00074318
29. Sariyar M, Borg A. Record Linkage Functions for Linking and Deduplicating Data Sets [Internet]. New York: Datacamp; 2022. [update 2022 Jan 9; cited 2022 May 2]. Available from: <https://cran.r-project.org/web/packages/RecordLinkage/RecordLinkage.pdf>

## ABSTRACT

**Objective:** To present a standardized methodology for linking different public health databases. **Methods:** This was a methodological review article specifically describing data processing procedures for deterministic linkage between structured databases. It instructs on how to: treat data, select linkage keys, and link databases using two databases simulated in R software. **Results:** The commands used for the deterministic linkage of the `inner_join` type were presented. The linkage process the resulted dropped to database with 40,108 pairs using only the "Name" key. Adding the second key, "Name of mother", the result dropped to 112 pairs. By adding the third key, "Date of birth", only two pairs were identified. **Conclusion:** Database linkage and its analysis are valid and valuable tools for health services in supporting health surveillance actions.

**Keywords:** Data Analysis; Epidemiology; Public Health; Public Health Surveillance.

## RESUMEN

**Objetivo:** Presentar metodología estandarizada para vincular diferentes bases de datos de salud pública. **Métodos:** Artículo de revisión metodológica y descripción de los procesos de tratamiento de datos para la vinculación determinista entre bases de datos. Se dieron instrucciones sobre como manejar los datos, seleccionar claves de vinculación y vincular las bases de datos empleando dos bases de datos simuladas en el software R. **Resultados:** Se presentaron los comandos utilizados para la vinculación determinista, del tipo `inner-join`. El proceso resultó en una base de datos con 40.108 pares utilizando únicamente la clave "Nombre". Con la adición de la segunda clave, "Nombre de la madre", el resultado se redujo a 112 pares. Al agregar la tercera clave, "Fecha de nacimiento", solo se identificaron dos pares. **Conclusión:** La vinculación de bases de datos y sus análisis son herramientas válidas y útiles para que los servicios de salud las utilicen para apoyar las acciones de vigilancia en la salud.

**Palabras clave:** Análisis de Datos; Epidemiología; Salud Pública; Vigilancia en Salud Pública.

Artigo publicado – Aspectos metodológicos sobre *record linkage*

Article

**Record Linkage for Malaria Deaths Data Recovery and Surveillance in Brazil**Klauss Kleydmann Sabino Garcia <sup>1,\*</sup>, Danielly Batista Xavier <sup>2</sup>, Seyi Soremekun <sup>3</sup>, Amanda Amaral Abrahão <sup>4</sup>, Chris Drakeley <sup>3</sup>, Walter Massa Ramalho <sup>1</sup> and André M. Siqueira <sup>5</sup><sup>1</sup> Center of Tropical Medicine, University of Brasilia, Brasilia 70904-970, Brazil<sup>2</sup> Escola Superior de Agricultura Luis de Queiroz, Universidade de São Paulo, Piracicaba 13418-900, Brazil<sup>3</sup> Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, University of London, London WC1E 7HT, UK<sup>4</sup> Faculty of Health Sciences, University of Brasilia, Brasilia 70910-900, Brazil<sup>5</sup> Oswaldo Cruz Foundation, Evandro Chagas National Institute of Infectology, Rio de Janeiro 21040-360, Brazil

\* Correspondence: kleydmann25@gmail.com

**Abstract:** Objective: The objective is to describe the results and the methodological processes of record linkage for matching deaths and malaria cases. Methods: A descriptive cross-sectional study was conducted with probabilistic record linkage of death and malaria cases data in Brazil from 2011 to 2020 using death records from the Mortality Information System (SIM) and epidemiological data from the Notifiable Diseases Information System (Sinan) and Epidemiological Surveillance Information Systems for malaria (Sivep-Malaria). Three matching keys were used: patient's name, date of birth, and mother's name, with an analysis of cosine and Levenshtein dissimilarity measures. Results: A total of 490 malaria deaths were recorded in Brazil between 2011 and 2020. The record linkage resulted in the pairing of 216 deaths (44.0%). Pairings where all three matching keys were identical accounted for 30.1% of the total matched deaths, 39.4% of the matched deaths had two identical variables, and 30.5% had only one of the three key variables identical. The distribution of the variables of the matched deaths (216) was similar to the distribution of all recorded deaths (490). Out of the 216 matched deaths, 80 (37.0%) had poorly specified causes of death in the SIM. Conclusions: The record linkage allowed for the detailing of the data with additional information from other epidemiological systems. Record linkage enables data linkage between information systems that lack interoperability and is an extremely useful tool for refining health situation analyses and improving malaria death surveillance in Brazil.

**Keywords:** malaria; epidemiology; public health; control; interrupted time series; Brazil

**Citation:** Garcia, K.K.S.; Xavier, D.B.; Soremekun, S.; Abrahão, A.A.; Drakeley, C.; Ramalho, W.M.; Siqueira, A.M. Record Linkage for Malaria Deaths Data Recovery and Surveillance in Brazil. *Trop. Med. Infect. Dis.* **2023**, *8*, 519. <https://doi.org/10.3390/tropicalmed8120519>

Academic Editor: Hans-Peter Fuehrer

Received: 2 November 2023

Revised: 4 December 2023

Accepted: 5 December 2023

Published: 14 December 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**1. Introduction**

Brazil's health surveillance system boasts a rich repository of epidemiological data stored in various Health Information Systems (HISs). Some of the HISs most frequently utilised by Health Surveillance include the Mortality Information System (SIM), housing information sourced from death certificates (DC); the Notifiable Diseases Information System (Sinan) and the Epidemiological Surveillance Information Systems (Sivep-Malária), both containing data from notification forms for diseases and notifiable conditions; the Hospital Information System (SIH), featuring data on hospital admissions; and the Live Birth Information System (SINASC), containing information on live births in Brazil [1].

The interoperability of these systems has not yet been achieved [1], and there is no single unique patient (user) identifier allowing the linkage between different databases. Historically, most records lack national individual identification numbers such as the Individual Taxpayer Registration (Cadastro de Pessoa Física—CPF) [2]. Consequently, when a need to correlate information across different databases arises, it becomes necessary to link data from the previously nominated HIS using probabilistic data linkage techniques.

Probabilistic linkage can be understood as a technique that considers a set of combination rules that analyse the probability of distinct records belonging to the same individual by calculating measures of similarity between open-field, categorical, date, or dissimilarity measures between numerical variables [3]. These techniques employ measures of similarity between records using variables such as name, mother's name, date of birth, gender, and place of residence, among others [4].

Probabilistic data linkage has yielded favourable results in research across various fields of knowledge and provided evidence for quality health outcomes within the public health sector [5–7]. However, the linkage not only depends on specialised human resources but also relies on the quality of data within these HIS records, and the availability of computational processing power for the task [2].

Given the necessity to integrate information from different HISs and facing limitations related to data quality and accessibility, probabilistic data linkage methods (record linkage) have proven to be a viable alternative for this integration [3]. This method has shown versatility by working with a set of rules that allow the analysis of similarity between records. The technique also enables the assessment of possible related pairs as true or false and maximises the number of true pairs identified [8]. Therefore, probabilistic linkage techniques present better results than deterministic record linkage because its sensitivity allows one to find more true pairs than the deterministic one, which matches only perfect records between databases [2,9].

The HISs used in the epidemiological surveillance of malaria are Sivep-Malaria and Sinan. Malaria cases reported in the Amazon region (an endemic area that covers the northern region of Brazil and the states Maranhão and Mato Grosso) are notified in Sivep-Malaria, while cases outside the Amazon region (area not endemic) are reported in Sinan. These individual notification forms do not collect information about case progression or mortality [10,11]. The Amazon region is responsible for approximately 99% of malaria cases in Brazil, while the other 1% that happens in the non-endemic area is majorly imported from the Amazon region or other countries. Because the occurrence of malaria in the non-endemic region is uncommon, diagnosis and treatment may occur late [12]. Possibly due to these factors that are related to surveillance and assistance matters, the non-endemic region can present fatality rates varying between 0.8% and 3.3% [13]. In 2019, the fatality rate in the non-endemic region was 123 times greater than the lethality rate in the Amazon region, which was under 0.02% [14].

For the official monitoring of malaria-related deaths, the National Malaria Control and Prevention Program (NMCP) utilises SIM data, which encompasses deaths caused by different malaria species. To investigate specific issues related to malaria death surveillance, it is necessary to analyse information from different HISs together. It is crucial to have clarity regarding what is related to malaria-related deaths, especially considering that this is a preventable health event [15]. This information can promote actions that might prevent these deaths.

Therefore, in light of the need to enhance malaria death surveillance activities, this study considered it important to operationalise the linkage, while reporting the methodological processes, and describing the information retrieved from different HISs. This will facilitate the discussion of the procedures conducted, challenges encountered, and actions that can be developed with the newly available data.

The aim of this study is to describe the results and methodological processes used in data linkage (record linkage) for malaria-related deaths and cases of the disease. It is expected that this material can serve as a model for other record linkage processes and help malaria death surveillance.

## 2. Materials and Methods

It is a descriptive cross-sectional study that reports the methodological processes of record linkage technique for the validation of the technique for retrieving sociodemographic and epidemiological data on malaria-related deaths. Procedures and results of the data

linkage method are described in this work. The data pertain to the total number of malaria-related deaths recorded in the SIM (ICD-10: B50, B51, B52, B53, and B54) and the total number of malaria-positive cases reported in Brazil in Sivep-Malaria and Sinan between 2011 and 2020.

From the SIM database, the following variables were utilised: date of death, date of birth, patient's name, mother's name, gender, age, race/ethnicity, education level, underlying cause of death, associated causes of death, place of death occurrence, medical care, death investigation, and occupation.

From the Sivep-Malaria and Sinan databases, the following variables were used: date of symptom onset, date of notification, date of birth, patient's name, mother's name, gender, age, race/ethnicity, education level, laboratory test results, and type of slide (new case slide or possible recurrence).

Descriptive statistics techniques were employed for data description. Confidence intervals for the relative frequencies of the studied variables were calculated to assess similarity between the distributions of paired data through record linkage compared to the original data. The chi-square test was applied to verify differences in variable distribution between the original records and the linked data. A significance level of 0.05 was adopted. When the *p*-value is greater than 0.05, it indicates no difference between the proportions of variables among the databases. In other words, the linked records in this scenario will exhibit a variable distribution similar to the distribution of the original database. The HIS records were also compared to identify possible inconsistencies between sociodemographic variables and *Plasmodium* species information. Additionally, the quantity of information marked as "unknown", "blank", or ambiguously defined as "unspecified malaria" was examined.

This work received favourable opinion number 51246121.0.0000.5558 from the Ethics Committee on Research of the Faculty of Medicine of the University of Brasília, and data access was obtained through an official request submitted to the Office of the Secretary of Health Surveillance and Environment of the Brazilian Ministry of Health (BMoH). Authorization to handle this data was granted by the Coordination-General of Information and Epidemiological Analysis (CGIAE) and the Coordination-General of Zoonoses and Vector-Borne Diseases Surveillance (CGZV). Data handling was conducted by CGIAE in a secure and monitored environment, following security protocols to prevent misuse or data leakage.

For the development of the record linkage R language and R Studio interface (version 4.2.1) were used [16]. The R script is available in Supplementary Material S1.

#### *Specific Linkage Procedures*

**Pre-planning:** Variables of interest from the SIM, Sivep-Malaria, and Sinan databases were defined. For pairing, "Patient's Name", "Date of Birth", and "Mother's Name" were selected as linkage key variables, following suggestions from previous studies [2,17,18].

**Data treatment:** To perform data treatment and linkages, the tidyverse, lubridate, stringr, stringi, fuzzyjoin, stringdist, abjutils, and Rcpp packages from R software (version 4.2.1) were used. Data preparation followed the data cleaning and treatment methods previously reported [2]. To reduce errors and inconsistencies, an initial standardisation of key variables was performed in all three databases. All characters in the "Patient's Name" and "Mother's Name" variables were converted to uppercase. Name connectors such as "E", "A", "DA", "DE", "DO", "DAS", "DES", and "DOS", accents, symbols, and spaces were removed. Texts resembling "observations" were also removed from the "Patient's Name" field, including numbers, punctuation marks, "KG" (kilogram), "PESO" (weight), "TEL" or "TL" or "FONE" (telephone), "OBS" (observation), and the words "unknown" or "unspecified". The "Patient's Name" variable was split into two new variables: "first name" and "last name" to increase the sensitivity of the pairing method.

All described procedures were applied to the SIM, Sivep-Malaria, and Sinan databases. Due to the high number of notifications in the Sivep-Malaria database, notifications

from this system were divided into biennial blocks (2011–2012, 2012–2013... 2018–2019, 2019–2020) to facilitate data processing by the software and capture epidemiological information prior to death. Considering that malaria illness and death are acute events [19], there is no perceived loss of pairs in this blocking process.

Parameters and rules for linkage: probabilistic linkage was performed using the “fuzzyjoin” package, which requires defining parameters for comparing information from different databases to operate effectively. In this package, distance measures need to be defined for nominal, numeric, and date variables. Two distance measures used in the pairings are Levenshtein [20] for names and cosine [21] for last names. For dates, these same distance techniques can be chosen, considering possible incorrect entries.

In statistics, a measure of distance refers to a quantitative calculation used to determine the similarity or dissimilarity between data points, vectors, or distributions. Levenshtein distance measures the minimum number of single-character edits required to change one string into the other and the cosine distance measures the cosine of the angle between two vectors (in this case: names) [22].

The linkage considered the following keys simultaneously: “first name”, “last name”, “date of birth”, and “mother’s name”. Levenshtein distance value used was 3 for the “first name” and “date of birth”, while cosine distance was set at 0.5 and used for “last names” and “mother’s name”.

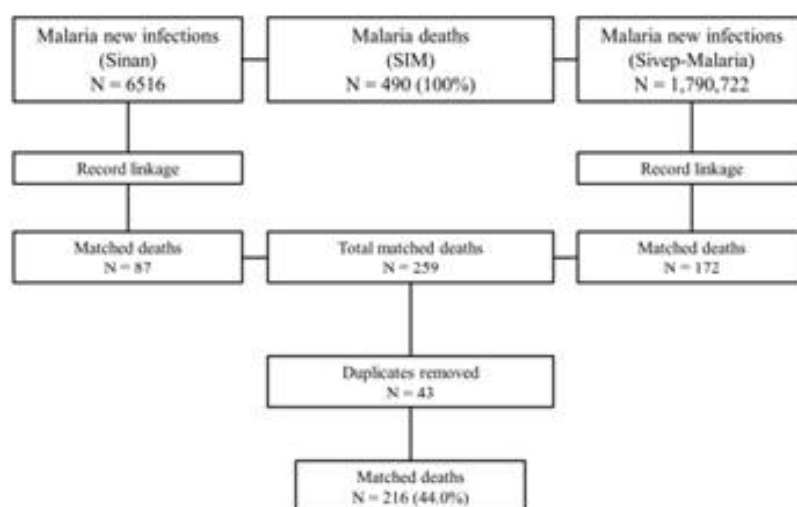
The value of 3 for Levenshtein distance allows up to 3 characters between records to be different and still to be considered a match, for example, “Danieli” and “Dannielly” and “1994-04-25” and “2094-04-26” (errors underlined) are possible to match. The use of value 2 in Levenshtein distance would make the parameter less sensitive. The Levenshtein distance value 3 (higher sensitivity than value 2) was used considering the common large amount of misspellings in first names and birth dates. The cosine distance can be set between 0 and 1, the value 0 indicates no similarity and 1 indicates perfect similarity. If the cosine value was set at 0.7, the search would be less sensitive.

Due to the impossibility of reviewing the identified pairs manually, after the linkage was concluded, only pairs that had at least 1 of the variable keys (last name, mother’s name, and date of birth) identical between records were considered true pairs. Those actions increased the specificity of the process. Therefore, the parameters were defined to have a moderate sensitivity and extremely high specificity.

Linkage execution: After defining the parameters and linkage rules, an inner join linkage was performed, where pairs from Sivep-Malaria and Sinan were linked to the SIM data. The “fuzzy\_inner\_join” function from the “fuzzyjoin” package [23] in R (version 4.1.1) was used for the linkage. Different linkage types can be reviewed in the methodological report by Garcia and colleagues (2022) [2].

### 3. Results

The SIM data totalled 490 deaths between 2011 and 2020, Sivep-Malaria presented 1,790,722 notifications, and Sinan summed 6516 records. After the SIM-Sinan linkage and the exclusion of records considered false positives by the criteria established in the study, 87 pairs were found. The SIM-Sivep linkage resulted in 172 pairs. In total, 259 pairs were found. Of the 259 pairs, 43 duplicates were identified (38 in the SIM-Sivep linkage and 5 in the SIM-Sinan linkage). This left 134 deaths from the SIM-Sivep linkage and 82 from the SIM-Sinan linkage, totalling 216 deaths with recovered information, equivalent to 44.0% of the 490 deaths registered during the study period (Figure 1).



**Figure 1.** Flowchart of the record linkage process between malaria deaths and cases.

Out of the 216 linked deaths, 148 (68.5%) had a perfect match for the “patient’s name” between the information systems, 117 (54.2%) had a perfect match for the “mother’s name”, and 166 (76.9%) records had a perfect match for the “date of birth”. Regarding the combination of key variables, 65 (30.1%) had a perfect match (identical patient’s name, mother’s name, and date of birth), 85 (39.4%) had a match with 2 identical variables, and 66 (30.5%) had only 1 identical variable (Table 1).

**Table 1.** Quantity of pairings by linkage keys used.

Concordance	Variables Used	Pairs	%
3 variables	Patient’s name	65	30.1%
	Mother’s name		
	Date of birth		
2 variables	Patient’s name	43	19.9%
	Mother’s name		
	Date of birth		
1 variable	Mother’s name	38	17.6%
	Date of birth		
Total		216	100.0%

The year with the highest number of matched deaths was 2012 with 32 pairs (48.4% of deaths registered in the year), and the year with the fewest pairs found was 2015, with only 12 pairs (31.5% of deaths in the year). On average, 49 deaths per year were recorded, of which approximately 21 were matched, meaning an average of 44.0% of annual deaths had epidemiological information supplemented from Sivep-Malaria and Sinan data.

The assessment of similarity between the paired data distributions ( $N = 216$ ) indicated that the proportions of the variables “gender”, “race/colour”, “basic cause of death”, “medical care”, “death investigation”, “year of death occurrence”, and “place of death occurrence” generally had distributions similar to the original database ( $N = 490$ ) (Table 2). The distribution of the variables “Age Group” and “Education” were the only ones where the distributions were not similar to the original database (Table 2).

Table 2. Profile of malaria mortality and paired deaths between 2011 and 2020, Brazil.

	Total Deaths	%	Paired Deaths	%	95% CI	p-Value (Chi-Squared)
	N = 490		N = 216			
Sex						0.8
Male	285	58.2%	128	59.3%	(52–66)	
Female	205	41.8%	88	40.7%	(34–48)	
Age Group						0.014
<1	24	4.9%	9	4.2%	(2.0–8.0)	
1–10	40	8.2%	21	9.7%	(6.3–15)	
11–20	38	7.8%	21	9.7%	(6.3–15)	
21–30	56	11.4%	29	13.4%	(9.3–19)	
31–40	70	14.3%	46	21.3%	(16–27)	
41–50	54	11%	32	14.8%	(10–20)	
51–60	63	12.9%	25	11.6%	(7.8–17)	
61–70	49	10%	14	6.5%	(3.7–11)	
71–80	47	9.6%	9	4.2%	(2.0–8.0)	
>80	48	9.80%	10	4.60%	(2.4–8.6)	
Blank	1	0.20%	-	-	-	
Race/Ethnicity						0.8
White	120	24.50%	53	24.50%	(19–31)	
Black	35	7.10%	14	6.50%	(3.7–11)	
Yellow	2	0.40%	-	-	-	
Brown	258	52.70%	112	51.90%	(45–59)	
Indigenous	65	13.30%	29	13.40%	(9.3–19)	
Blank	10	2%	8	3.70%	(1.7–7.4)	
Years of Education						0.004 *
None	86	17.60%	18	8.30%	(5.2–13)	
1–3 years	70	14.30%	24	11.10%	(7.4–16)	
4–7 years	74	15.10%	32	14.80%	(10–20)	
8–11 years	76	15.50%	44	20.40%	(15–26)	
12 years or more	56	11.40%	42	19.40%	(15–25)	
Ignored	48	9.80%	19	8.80%	(5.5–14)	
Blank	80	16.30%	37	17.10%	(12–23)	
Cause of death (ICD-10)						0.2
B50– <i>Plasmodium falciparum</i> (Pf)	101	20.6%	54	25%	(19–31)	
B51– <i>Plasmodium vivax</i> (Pv)	183	37.3%	79	36.6%	(30–43)	
B52– <i>Plasmodium malariae</i> (Pm)	5	1%	3	1.4%	(0.36–4.3)	
B53–Other Forms	6	1.2%	5	2.3%	(0.86–5.6)	
B54–Not Specified	151	30.8%	66	30.6%	(25–37)	
Malaria-associated cause	44	9%	9	4.2%	(2.0–8.0)	
Received medical care?						0.4
Yes	311	63.5%	150	69.4%	(63–75)	
No	49	10%	17	7.9%	(4.8–13)	
Unknown	21	4.3%	10	4.6%	(2.4–8.6)	
Blank	109	22.2%	39	18.1%	(13–24)	
Was the death investigated?						0.3
Yes	94	19.2%	51	23.6%	(18–30)	
No	236	48.2%	94	43.5%	(37–50)	
Blank	160	32.7%	71	32.9%	(27–40)	
Year of Occurrence						>0.9
2011	72	14.7%	29	13.4%	(9.3–19)	
2012	66	13.5%	32	14.8%	(10–20)	
2013	45	9.2%	19	8.8%	(5.5–14)	
2014	38	7.8%	20	9.3%	(5.9–14)	
2015	38	7.8%	12	5.6%	(3.0–9.7)	
2016	39	8%	19	8.8%	(5.5–14)	
2017	39	8%	15	6.9%	(4.1–11)	

Table 2. Cont.

	Total Deaths	%	Paired Deaths	%	95% CI	p-Value (Chi-Squared)
2018	61	12.4%	30	13.9%	(9.7–19)	0.9
2019	39	8%	14	6.5%	(3.7–11)	
2020	53	10.8%	26	12%	(8.2–17)	
State of Occurrence						
Rondônia	37	7.6%	14	6.5%	(3.7–11)	
Acre	22	4.5%	3	1.4%	(0.36–4.3)	
Amazonas	97	19.8%	40	18.5%	(14–24)	
Roraima	51	10.4%	17	7.9%	(4.8–13)	
Pará	73	14.9%	28	13%	(8.9–18)	
Amapá	31	6.3%	18	8.3%	(5.2–13)	
Tocantins	2	0.4%	1	0.5%	(0.02–3.0)	
Maranhão	20	4.1%	4	1.9%	(0.59–5.0)	
Piauí	6	1.2%	4	1.9%	(0.59–5.0)	
Ceará	3	0.6%	1	0.5%	(0.02–3.0)	
Rio Grande do Norte	3	0.6%	1	0.5%	(0.02–3.0)	
Paraíba	2	0.4%	-	-	-	
Pernambuco	7	1.4%	2	0.9%	(0.16–3.7)	
Alagoas	1	0.2%	1	0.5%	(0.02–3.0)	
Sergipe	1	0.2%	1	0.5%	(0.02–3.0)	
Bahia	11	2.2%	6	2.8%	(1.1–6.2)	
Minas Gerais	25	5.1%	13	6%	(3.4–10)	
Espirito Santo	5	1%	3	1.4%	(0.36–4.3)	
Rio de Janeiro	9	1.8%	7	3.2%	(1.4–6.8)	
São Paulo	33	6.7%	23	10.6%	(7.0–16)	
Paraná	5	1%	3	1.4%	(0.36–4.3)	
Santa Catarina	6	1.2%	3	1.4%	(0.36–4.3)	
Rio Grande do Sul	5	1%	3	1.4%	(0.36–4.3)	
Mato Grosso do Sul	2	0.4%	1	0.5%	(0.02–3.0)	
Mato Grosso	16	3.3%	8	3.7%	(1.7–7.4)	
Goiás	10	2.0%	5	2.3%	(0.86–5.6)	
Distrito Federal	7	1.4%	6	2.8%	(1.1–6.2)	

Note: \* p-values less than 0.05 are considered statistically significant.

Despite the data, in general, having similar distributions, some variables did not have a similar distribution between the linked database and the original databases: there was an expected 14.3% of deaths between 31 and 40 years, but 21.3% were matched (95% CI: 16–27%), indicating more matches between 31 and 40 years than expected. The opposite occurred in the age groups of 71 to 80 years and over 80 years, which had fewer matches than expected (Table 2).

Dissimilar distributions were also found for people with no education; 8.3% were matched (95% CI: 5.2–13.0%) of the deaths, while the expected was 17.6%. For people with more than 11 years of education, 19.4% were matched (95% CI: 15.0–25.0%), and the expected was 11.4%.

Although the chi-square test indicated no difference between the groups, when analysing the variable "Cause of death", it is observed that there was less matching for deaths with malaria as an associated cause (4.2%; 95% CI: 2.0–8.0%) than expected (9.0%). The same was seen in the variable "State of occurrence", where fewer deaths were matched (1.4%; 95% CI: 0.3–4.3%) than expected (4.5%) in the state of Acre, and more matched deaths were found in São Paulo (10.6%) than expected (6.7%).

The causes of death described in the SIM showed some consistent data and some inconsistencies when compared with the data recovered from the Sivep-Malaria and Sinan information. Deaths from *Plasmodium falciparum* (N = 54; ICD-10: B50) had 20% (N = 11) inconsistency with epidemiological data, which indicated infections by *Plasmodium vivax* instead of *falciparum*. Deaths from *Plasmodium vivax* (N = 79; B51) had only 5.0% (N = 4)

inconsistencies detected. Of the three deaths from *Plasmodium malariae* (B52), two were marked as infections by *Plasmodium falciparum*. The five deaths from “other forms of malaria” (B53) presented positive tests for *Plasmodium falciparum* (N = 1), *Plasmodium vivax* (N = 3), and mixed infections (N = 1). For deaths marked as “unspecified malaria” (N = 66), 39% (N = 26) were infections by *falciparum*, 48% (N = 32) were *vivax* infections, 7.6% (N = 5) were mixed infections, one (1.5%) was caused by *Plasmodium malariae*, one (1.5%) by *Plasmodium ovale*, and one (1.5%) had a result of “non-*falciparum*”. Deaths where malaria was not the underlying cause (N = 9) were seven (78%) *P. vivax* infections, one (11%) *P. falciparum* infection, and one (11%) mixed infection (Table 3).

**Table 3.** Basic causes of death (SIM) and laboratory results recovery (Sivep-Malaria and Sinan).

	B50—Pf	B51—Pv	B52—Pm	B53—Other Forms of Malaria	B54— Unspecified Malaria	Malaria— Associated Cause	Total
<i>Pf</i>	39 (72.2%)	2 (2.5%)	2 (66.7%)	1 (20.0%)	26 (39.4%)	1 (11.1%)	71 (32.9%)
<i>Pv</i>	10 (18.5%)	75 (94.9%)	0 (0.0%)	3 (60.0%)	32 (48.5%)	7 (77.8%)	127 (58.8%)
Mixed	4 (7.4%)	1 (1.3%)	0 (0.0%)	1 (20.0%)	5 (7.6%)	1 (11.1%)	12 (5.6%)
<i>Pm</i>	0 (0.0%)	0 (0.0%)	1 (33.3%)	0 (0.0%)	1 (1.5%)	0 (0.0%)	2 (0.9%)
<i>P. ovale</i>	1 (1.9%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.5%)	0 (0.0%)	2 (0.9%)
Non- <i>Pf</i>	0 (0.0%)	1 (1.3%)	0 (0.0%)	0 (0.0%)	1 (1.5%)	0 (0.0%)	2 (0.9%)
Total	54 (25.0%)	79 (36.6%)	3 (1.4%)	5 (2.3%)	66 (30.6%)	9 (4.2%)	216 (100.0%)

Regarding the recovery of data for “unknown” or “blank” entries for other variables from the information systems, for the “education” field, 37 (75.5%) records were recovered out of 49 marked as “unknown” or “blank” in the information systems (Table S1); for “race/color”, 8 (72.7%) out of 11 “unknown” or “blank” entries in Sivep-Malaria or Sinan were recovered (Table S2); for the “occupation” field in the SIM and “activity in the last 15 days”, 56 records (60.2%) were recovered from SIM occupation compared to 93 entries marked as “unknown” or “blank” in the “activity in the last 15 days” variable in Sivep-Malaria or Sinan (Table S3).

#### 4. Discussion

Approximately 30.6% of information regarding the malaria species causing deaths was not specified in the death certificate. However, record linkage allowed the identification of this information, estimating that almost half of the deaths due to unspecified malaria were caused by *Plasmodium vivax* infections, and nearly 40% were due to *Plasmodium falciparum* infections.

National, state, and municipal malaria death surveillance can use this information to direct disease control actions to prevent deaths. With record linkage, not only can the cause of death be determined, but also the previously nonspecific basic causes can be detailed, enabling more precise information on the locations of deaths (registered in SIM) and the locations of likely infections (registered in Sivep-Malaria and Sinan).

There was a balance between the distributions of cases across variables, indicating that the record linkage technique was able to identify true pair records with similar variable distributions to the original database. That indicates that epidemiological analysis could be carried out using the matched pairs and that those pairs would have a similar distribution to the original dataset.

The combination of information from these HISs allows surveillance services to expand and improve death investigation, enabling the identification of more specific health situations. Initially, the estimate was that about 25.0% of paired deaths were due to *Plasmodium falciparum* and 36.6% to *Plasmodium vivax*. However, with the data recovery, the estimate changes to 37% and 51.4%, respectively. This highlights a weakness in classifying the cause of death in SIM and raises concerns that *Plasmodium falciparum* deaths may be underestimated by almost 50% in Brazil.

The record linkage also allowed the recovery of sociodemographic data related to education level, race/ethnicity, and occupation/economic activities of patients. These pieces of information can be used by epidemiological surveillance services to refine mortality profile analyses and thus direct prevention actions with greater precision.

On the other hand, inconsistencies in the data that were also observed need to be investigated. It is clear that the consistency of information on gender, age group, and race/ethnicity was high (above 92%) because they are basic mandatory filling information on forms. However, information about education showed more than half of the records matched with inconsistencies between SIM data and Sivep-Malaria or Sinan data. Once again, this is information of high importance for supplementing the definition of the mortality profile and health situation analyses to enhance actions aimed at vulnerable groups.

Therefore, it is important for epidemiological surveillance departments to investigate and understand the reasons for these inconsistencies. These inconsistencies may be related to how data are collected during case notification or when filling out the death certificate, and this investigation may reveal other difficulties and limitations in the data collection process.

Furthermore, it is necessary to investigate the unpaired death records in SIM. Due to the sensitivity of this information, this task could not be conducted in this study. It is therefore suggested that the NMCP examine this issue to identify deaths that may not have been reported in the Sivep-Malaria or in the Sinan.

Moreover, as the malaria information systems do not provide specific information on deaths due to the disease, the same applies to the registration of its hospitalisations. Monitoring and investigating factors associated with malaria-related hospitalisations can help prevent such occurrences, thereby potentially reducing deaths. Therefore, it is recommended that this data linkage process also be carried out between the Hospital Information System, Sivep-Malaria, and Sinan.

Additionally, the incorporation of an "outcome" field in the individual notification form of Sivep-Malaria and Sinan should be evaluated in the NMCP. This would enable the tracking of the patient's progress and the collection of information regarding recovery, disease recurrence, transfers (between health facilities), treatment abandonment, hospitalisations, or death.

Another issue is the lack of a unique identifying variable across different databases. Currently, the HISs are designed with data models that do not have identification documents, such as the CPF, as mandatory record variables in the HIS [2]. The use of record linkage techniques and the undertaking of this work are motivated by the lack of interoperability between the different HISs inside the BMoH. However, the current epidemiological scenarios urge the integration of these different pieces of information, given the difficulty of working with data from various sources [1].

It is important for the IT department responsible for these systems, in conjunction with the health departments, to develop projects and/or updates for the HIS to make this data reliable and make it available to the federated entities. This would enable more qualified analyses of the population's health status.

Regarding the methods employed in this study, the study achieved a 44.0% data recovery rate out of the total deaths. This indicates low data quality in the systems, particularly with respect to the filling of personal information. Poor data quality has previously been reported as a factor that has a negative influence on probabilistic linkage results [24,25]. This highlights the need to not only train epidemiological surveillance units and technicians but also healthcare professionals responsible for registering death certificates.

Table 1 shows that the use of methods that only allow pairing of records with a perfect match between them results in utilising only 30.1% of the total matched deaths. This highlights that the use of probabilistic methods for data linkage increases the number of linked pairs by at least two times.

The keys used in the linkage process showed a higher frequency of records with "date of birth" correctly filled out between the databases. This may indicate a similar scenario

in other databases. Furthermore, among pairs that had only one identical key variable between the systems, “date of birth” allowed the highest number of pairings when the information from the other two key variables was dissimilar (38 pairs; 17.6%) between the systems.

Therefore, it is strongly recommended that the “date of birth” variable be frequently used in probabilistic data linkage processes, as it tends to have fewer filling errors compared to the “patient’s name” and “mother’s name” variables.

Since the linkage rules consider possible spelling errors in filling out key variables for pairing, death records may not have been matched due to an excess of spelling errors in the Sivep-Malaria and Sinan forms. Hence, the low pairing of records for individuals over 71 years of age may be related to lower educational levels. Conversely, among users aged 31 to 40, which had higher pairing rates, the predominant education level is above 8 years of study (Tables S4–S6). The same pattern can be observed for deaths recorded with the basic cause “Malaria as an associated cause” (Table S7), as most of these records are of users with “illiterate” schooling.

The higher or lower pairing than expected for some variables is related to the education level of the deceased, and it is known that elderly people in Brazil have lower levels of education [26]. Therefore, in addition to highlighting that data quality is a critical point to optimise the data linkage process between different HISs, it is also emphasised that epidemiological surveillance technicians should pay closer attention when collecting information from individuals with lower education levels or those in older age groups (elderly).

It is important to note that the definition of rules and parameters for linkage can be adjusted according to the level of data quality to be used. The use of Levenshtein distances [20] and cosine similarity [21] can be calibrated to increase the number of pairs found, but this increase may necessitate a manual review of the pairs to exclude false pairs. Depending on the magnitude of the results, this manual review may become impractical [27]. Also, alternative methods can be investigated, such as the use of deterministic linking methods [2,9] or other packages and functions available for the R software, such as *fastlink* [28], or even the utilisation of other software like *Reclink* [29] or *Python* [30].

In 2018, the General Data Protection Law (LGPD) (Law No. 13,709, of 14 August 2018) [31] was enacted, which emphasises access and the treatment of personal data. This law introduced new rules for accessing information and stricter rules for accessing nominal data for research purposes. LGPD advocates that data provided for research purposes be anonymised whenever possible. For public health research, LGPD allows the sharing of personal data as long as they are exclusively processed within the organisation and strictly for the purpose of conducting studies, these processes can make the research execution process longer than initially planned.

Data linkages have enabled the identification of new information for health surveillance, supporting data quality assessment processes, information supplementation, and the development of public health policies. This information allows for understanding the profile of users in the healthcare field, individual analyses, cohort tracking, and understanding the paths taken by populations through healthcare services.

Finally, this study presents an alternative for the data linkage of the Information Systems in the SUS through free and open statistical and computational techniques. These techniques may allow researchers to analyse data from the same individual across different systems, aggregating a larger number of variables. This enables the identification of morbidity and mortality patterns, which can support the development of public policies that promote and prevent the occurrence of deaths.

Considering that the HISs used in this study are also used by different sectors inside the BMoH [1] it is possible to say that the methods presented are applicable to the surveillance of other death causes such as dengue, HIV/AIDS, tuberculosis, etc. As well, probabilistic linkage techniques to investigate these causes of death have already been reported, but using *Reclink* software [32–34].

The limitations of this study include potential systematic errors arising from the use of secondary databases, such as poor quality of personal data, which may have impacted the total number of pairs identified. Additionally, it is possible that some true pairs were excluded due to the automatic classification of true pairs being the ones with at least one identical key variable among records.

## 5. Conclusions

Probabilistic data linkage techniques enable the relationship of data that lacks a unique identification key. In a scenario where different HISs lack identification variables for users, a probabilistic linkage is a valuable tool for better defining the mortality profile, refining health situation analyses, and increasing the precision of information related to causes of death.

The recovery of epidemiological information on malaria cases for investigating deaths due to the disease is essential to support the National Malaria Control and Prevention Program (NMCP) in monitoring malaria-related deaths and controlling severe malaria cases that may result in death.

It is important that protocols for the acquisition or management of sensitive data are better described at the federal, state, and municipal levels to enable data linkage processes to be carried out by educational and research institutions in addition to the technical units of healthcare services. This collaboration between academia and healthcare services is essential for strengthening public health in the country.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/tropicalmed8120519/s1>, Supplementary file S1: R script for record linkage; Supplementary file S2: Table S1. Retrieval of "ignored" or "blank" data for the education level variable (in years of study) in SIM, Table S2. Recovery of "ignored" or "blank" data for the race/colour variable in SIM, Table S3. Recovery of "ignored" or "blank" data for the occupation variable in SIM, Table S4. Percentage of matching deaths among people with the underlying cause of death as "Malaria as an associated cause" according to age group in years, Table S5. Percentage of matching deaths of people between 31 and 40 years old according to level of education in years of study, Table S6. Percentage of matching deaths of people over 71 years of age according to level of education in years of study, Table S7. Percentage of matching deaths among people with the underlying cause of death as "Malaria as an associated cause" according to level of education in years of study.

**Author Contributions:** K.K.S.G. is the principal investigator responsible for the overall conduct of the study. K.K.S.G., D.B.X., A.A.A., W.M.R., S.S., C.D. and A.M.S. conceptualised the study. K.K.S.G. and D.B.X. were responsible for data curation and accessed and verified all the data in the study. K.K.S.G. performed the statistical analysis. K.K.S.G. wrote the manuscript with support from A.A.A. and A.M.S. All authors reviewed and edited the manuscript. All authors had full access to all the data in the study, have read and approved the final version of the manuscript, and accept responsibility for the decision to submit for publication. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work received funding from the Bill & Melinda Gates Foundation (INV-003970), the Brazilian Ministry of Health (MoH/DECIT/CNPq—#443148/2019–8), and the Higher Education Personnel Improvement Coordination notice 10/2022 (CAPES—#88881.690401/2022-01).

**Institutional Review Board Statement:** This work was approved by the Ethics Committee on Research of the Faculty of Medicine of the University of Brasília in the process number 51246121.0.0000.5558.

**Informed Consent Statement:** Patient consent was waived because nominal data was not used by the authors and all sensitive data was handled only by the Ministry of Health of Brazil. Those procedures are in accordance with the resolution number 466/2012 from the National Health Council (Chapter IV, paragraph VII) and the General Personal Data Protection Law (LGPD) of Brazil, Law No. 13709, of 14 August 2018.

**Data Availability Statement:** Sensitive information such as names and dates of birth are not available. Other data are available upon reasonable request.

**Acknowledgments:** Thanks to all authors' contributions. To CAPES (Higher Education Personnel Improvement Coordination) for financing the sandwich doctorate between the Nucleus of Tropical Medicine from the University of Brasília and The London School of Hygiene and Tropical Medicine. Thanks to both institutions for providing knowledge and means to develop the research.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

- Coelho, G.C.; Chioro, A. Afinal, quantos Sistemas de Informação em Saúde de base nacional existem no Brasil? *Cad. Saúde Pública* **2021**, *37*, e00182119. [CrossRef] [PubMed]
- Garcia, K.K.S.; de Miranda, C.B.; de Sousa, F.N.E.F. Procedures for health data linkage: Applications in health surveillance. *Epidemiol. Serv. Saúde* **2022**, *31*, e20211272. [CrossRef] [PubMed]
- Avoundjian, T.; Dombrowski, J.C.; Golden, M.R.; Hughes, J.P.; Guthrie, B.L.; Baseman, J.; Sadinle, M. Comparing Methods for Record Linkage for Public Health Action: Matching Algorithm Validation Study. *JMIR Public Health Surveill.* **2020**, *6*, e15917. [CrossRef] [PubMed]
- Enamorado, T.; Imai, K. Validating Self-Reported Turnout by Linking Public Opinion Surveys with Administrative Records. *Public Opin. Q.* **2019**, *83*, 723–748. [CrossRef]
- Maia, L.T.S.; Souza, W.V.; Mendes, A.C.G. A contribuição do linkage entre o SIM e SINASC para a melhoria das informações da mortalidade infantil em cinco cidades brasileiras. *Rev. Bras. Saúde Matern. Infant.* **2015**, *15*, 57–66. [CrossRef]
- Rocha, M.S.; Oliveira, G.P.; Guillen, L.C.T.; Coeli, C.M.; Saraceni, V.; Pinheiro, R.S. Uso de linkage entre diferentes bases de dados para qualificação de variáveis do Sinan-TB e a partir de regras de scripting. *Cad. Saúde Pública* **2019**, *35*, e00074318. [CrossRef]
- Shah, G.H.; Lertwachara, K.; Ayanso, A. Record linkage in healthcare: Applications, opportunities, and challenges for public health. *Int. J. Healthc. Deliv. Reform Initiat.* **2010**, *2*, 29–47. [CrossRef]
- Sayers, A.; Ben-Shlomo, Y.; Blom, A.W.; Steele, F. Probabilistic record linkage. *Int. J. Epidemiol.* **2016**, *45*, 954–964. [CrossRef]
- Zhu, Y.; Matsuyama, Y.; Ohashi, Y.; Setoguchi, S. When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. *J. Biomed. Inform.* **2015**, *56*, 80–86. [CrossRef] [PubMed]
- Brasil. Ministério da Saúde. Ficha de Notificação Sivep-Malária Atualizado em 30 de jun. 2022. Available online: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/publicacoes-svs/malaria/ficha-de-notificacao-sivep-malaria-v03-08-2020.pdf/@download/file/Ficha%20de%20Notifica%C3%A7%C3%A3o%20Sivep-Mal%C3%A1ria%20v03-08-2020.pdf> (accessed on 31 August 2022).
- Brasil. Ministério da Saúde. Portal Sinan–Malária: Ficha de Notificação/Investigação Individual Para Malária. Available online: [http://portalsinan.saude.gov.br/images/documentos/Agravos/Malaria/Malaria\\_v5.pdf](http://portalsinan.saude.gov.br/images/documentos/Agravos/Malaria/Malaria_v5.pdf) (accessed on 31 August 2022).
- Lorenz, C.; Virginio, F.; Aguiar, B.S.; Suesdek, L.; Chiaravalloti-Neto, F. Spatial and temporal epidemiology of malaria in extra-Amazonian regions of Brazil. *Malar J.* **2015**, *14*, 408. [CrossRef] [PubMed]
- Garcia, K.K.S.; Abrahão, A.A.; Oliveira, A.F.M.; Henriques, K.M.D.; de Pina-Costa, A.; Siqueira, A.M.; Ramalho, W.M. Malaria time series in the extra-Amazon region of Brazil: Epidemiological scenario and a two-year prediction model. *Malaria J.* **2022**, *21*, 157. [CrossRef] [PubMed]
- Brazilian Ministry of Health. Secretariat of Health Surveillance, Department of Immunization and Communicable Diseases, General Coordination of Surveillance of Zoonoses and Vector-Transmitted Diseases. Epidemiological Bulletin–Malaria: 2020. November 2020; p. 118. Available online: [https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/boletins-epidemiologicos/especiais/2020/boletim\\_especial\\_malaria\\_1dez20\\_final.pdf](https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/boletins-epidemiologicos/especiais/2020/boletim_especial_malaria_1dez20_final.pdf) (accessed on 31 August 2022).
- Malta, D.C.; Duarte, E.C. Causas de mortes evitáveis por ações efetivas dos serviços de saúde: Uma revisão da literatura. *Ciênc Saúde Coletiva* **2007**, *12*, 765–776. [CrossRef] [PubMed]
- R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023; Available online: <https://www.R-project.org/> (accessed on 31 August 2022).
- Churches, T.; Christen, P.; Lim, K.; Zhu, J.X. Preparation of name and address data for record linkage using hidden Markov models. *BMC Med. Inform. Decis. Mak.* **2002**, *2*, 9. [CrossRef] [PubMed]
- Yaohao, P.; Mation, L.F. *O Desafio do Pareamento de Grandes Bases de Dados: Mapeamento de Métodos de Record Linkage Probabilístico e Diagnóstico de Sua Viabilidade Empírica*; Instituto de Pesquisa Econômica Aplicada (IPEA): Rio de Janeiro, Brazil, 2018; ISSN 1415-4765.
- Organização Mundial da Saúde (OMS). Malaria. Update at 26 July 2022. Available online: <https://www.who.int/news-room/fact-sheets/detail/malaria#:~:text=Left%20untreated,%20P,a%20period%20of%2024%20hours> (accessed on 5 September 2022).
- Zhang, S.; Hu, Y.; Bian, G. Research on string similarity algorithm based on Levenshtein Distance. In Proceedings of the 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 25–26 March 2017; pp. 2247–2251.
- Liao, H.; Xu, Z. Approaches to manage hesitant fuzzy linguistic information based on the cosine distance and similarity measures for HFLTSs and their application in qualitative decision making. *Expert Syst. Appl.* **2015**, *42*, 5328–5336. [CrossRef]

22. Guo, H.; Li, Y.; Liu, Y.; Li, W.; Li, X. Evaluation of string comparators for record linkage in Chinese environment. *Int. J. Wavelets Multiresolut. Inf. Process.* **2022**, *20*, 2250023. [CrossRef]
23. David, R.; Jennifer, B.; Joran, E. FuzzyJoin. New York: Datacamp. 2020 [Update 15 May 2020]. Available online: <https://cran.r-project.org/web/packages/fuzzyjoin/fuzzyjoin.pdf> (accessed on 2 September 2022).
24. Dusetzina, S.B.; Tyree, S.; Meyer, A.M.; Meyer, A.; Green, L.; Carpenter, W.R. *Linking Data for Health Services Research: A Framework and Instructional Guide*; Agency for Healthcare Research and Quality (US): Rockville, MD, USA, 2014.
25. Coeli, C.M.; Saraceni, V.; Medeiros, P.M., Jr.; da Silva Santos, H.P.; Guillen, L.C.T.; Alves, L.G.S.B.; Hone, T.; Millett, C.; Trajman, A.; Durovni, B. Record linkage under suboptimal conditions for data-intensive evaluation of primary care in Rio de Janeiro, Brazil. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 190. [CrossRef]
26. Castro, C.M.S.; Costa, M.F.L.; Cesar, C.C.; Neves, J.A.B.; Sampaio, R.F. Influência da escolaridade e das condições de saúde no trabalho remunerado de idosos brasileiros. *Cien. Saude Colet.* **2019**, *24*, 4153–4162. [CrossRef]
27. Grannis, S.J.; Overhage, J.M.; Hui, S.; McDonald, C.J. Analysis of a probabilistic record linkage technique without human review. *AMIA Annu. Symp. Proc.* **2003**, *2003*, 259–263.
28. Enamorado, T.; Fifield, B.; Imai, K. FastLink. New York: Datacamp. 2020 [Update 29 April 2020]. Available online: <https://cran.r-project.org/web/packages/fastLink/fastLink.pdf> (accessed on 2 September 2022).
29. Camarg, K.R.d., Jr.; Coeli, C.M. Reclink: Aplicativo para o relacionamento de bases de dados, implementando o método probabilistic record linkage. *Cad. Saúde Pública* **2000**, *16*, 439–447. [CrossRef]
30. Nahorniak, J.; Bovbjerg, V.; Case, S.; Kincl, L. Application of data linkage techniques to Pacific Northwest commercial fishing injury and fatality data. *Inj. Epidemiol.* **2021**, *8*, 26. [CrossRef]
31. Brasil. Presidência da República. Secretaria-Geral. Subchefia para Assuntos Jurídicos. Lei Geral de Proteção de Dados Pessoais (LGPD)-LEI N° 13.709, DE 14 DE AGOSTO DE 2018. Available online: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/L13709compilado.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm) (accessed on 5 September 2022).
32. Coelho, G.E.; Leal, P.L.; Cerroni, M.D.P.; Simplicio, A.C.R.; Siqueira, J.B., Jr. Sensitivity of the Dengue Surveillance System in Brazil for Detecting Hospitalized Cases. *PLoS Negl. Trop. Dis.* **2016**, *10*, e0004705. [CrossRef]
33. Fonseca, M.G.P.; Coeli, C.M.; Lucena, F.D.F.D.A.; Veloso, V.G.; Carvalho, M.S. Accuracy of a probabilistic record linkage strategy applied to identify deaths among cases reported to the Brazilian AIDS surveillance database. *Cad. Saúde Pública* **2010**, *26*, 1431–1438. [CrossRef]
34. Oliveira, G.P.D.; Bierrenbach, A.L.D.S.; Camargo Júnior, K.R.D.; Coeli, C.M.; Pinheiro, R.S. Accuracy of probabilistic and deterministic record linkage: The case of tuberculosis. *Rev. Saude Publica* **2016**, *50*. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

