

APLICAÇÕES DA BIOESTATÍSTICA NA VIGILÂNCIA DA MALÁRIA NA AMAZÔNIA BRASILEIRA

Vinícius Aguiar Alcântara da Silva¹.

¹Universidade do Estado do Pará (UEPA), Belém, Pará.

<http://lattes.cnpq.br/0491463969707287>

RESUMO: A malária apresenta distribuição territorial heterogênea na Amazônia brasileira e exige métodos capazes de distinguir variações aleatórias de mudanças epidemiologicamente relevantes. Este capítulo discute como a bioestatística transforma registros de vigilância em evidências para planejamento, previsão e avaliação das ações de controle. Realizou-se revisão narrativa e documental de artigos e documentos técnicos publicados, preferencialmente, entre 2020 e 2026. Foram sistematizadas aplicações de estatística descritiva, indicadores de incidência, séries temporais, regressão para dados de contagem, análise espacial, modelos hierárquicos, aprendizado de máquina e métricas de validação. Os estudos demonstram que médias e frequências são indispensáveis, porém insuficientes para explicar sazonalidade, autocorrelação, excesso de zeros e heterogeneidade entre municípios, áreas indígenas, rurais e de garimpo. Modelos temporais e espaço-temporais permitem identificar tendências, quantificar associações ambientais e produzir previsões; entretanto, resultados dependem da qualidade do SIVEP-Malária, da escolha adequada dos denominadores e da avaliação de incertezas. Conclui-se que o uso responsável da bioestatística fortalece a vigilância preditiva e a alocação territorial de recursos, contribuindo para as metas de eliminação da doença.

PALAVRAS-CHAVE: Modelagem preditiva. Séries temporais. Saúde pública.

APPLICATIONS OF BIOSTATISTICS IN MALARIA SURVEILLANCE IN THE BRAZILIAN AMAZON

ABSTRACT: Malaria has a heterogeneous territorial distribution in the Brazilian Amazon and requires methods capable of distinguishing random variation from epidemiologically relevant changes. This chapter discusses how biostatistics turns surveillance records into evidence for planning, forecasting, and evaluating control actions. A narrative and documentary review of articles and technical documents, preferably published between 2020 and 2026, was conducted. Applications of descriptive statistics, incidence indicators, time-series analysis, regression for count data, spatial analysis, hierarchical models, machine learning,

and validation metrics were systematized. The studies show that averages and frequencies are essential but insufficient to explain seasonality, autocorrelation, excess zeros, and heterogeneity among municipalities, Indigenous areas, rural territories, and mining zones. Temporal and spatiotemporal models can identify trends, quantify environmental associations, and generate forecasts; however, results depend on SIVEP-Malaria data quality, appropriate denominators, and uncertainty assessment. Responsible use of biostatistics strengthens predictive surveillance and territorial resource allocation, supporting malaria elimination goals.

KEYWORDS: Predictive modeling. Time series. Public health.

INTRODUÇÃO

A malária é uma doença parasitária transmitida por mosquitos do gênero *Anopheles* e causada por protozoários do gênero *Plasmodium*. No Brasil, a transmissão concentra-se na Amazônia, onde extensas distâncias, mobilidade populacional, diversidade ambiental e desigualdades no acesso ao diagnóstico produzem padrões epidemiológicos complexos. A predominância de *Plasmodium vivax*, a persistência de focos por *P. falciparum* e a ocorrência de transmissão em áreas indígenas, rurais, urbanas e de garimpo exigem estratégias diferenciadas de vigilância (BRASIL, 2024a).

O Sistema de Informação de Vigilância Epidemiológica da Malária reúne registros empregados no acompanhamento da região amazônica. A disponibilidade de datas, localidade, espécie parasitária, características demográficas, diagnóstico e tratamento permite construir indicadores e avaliar tendências. Entretanto, grandes bases de dados não geram conhecimento automaticamente. É necessário selecionar medidas, verificar pressupostos, controlar vieses, representar a incerteza e traduzir resultados para decisões sanitárias.

A bioestatística fornece métodos para descrever a ocorrência da doença, testar hipóteses, estimar associações e prever cenários. Em vigilância, ela auxilia a responder perguntas práticas: houve aumento além do esperado? Quais municípios apresentam maior risco? A sazonalidade mudou? O desmatamento ou a mobilidade estão associados aos casos? Uma intervenção reduziu a incidência? Quantos testes e medicamentos serão necessários nas próximas semanas? Cada questão demanda desenho analítico e método compatíveis.

Os desafios metodológicos são expressivos. Casos de malária são eventos de contagem, frequentemente assimétricos, com variância superior à média e grande quantidade de zeros em áreas de baixa transmissão. Observações de semanas sucessivas não são independentes, e municípios próximos ou conectados compartilham fatores ambientais e fluxos humanos. Modelos lineares simples podem produzir estimativas inadequadas quando ignoram sobredispersão, autocorrelação sazonalidade e dependência espacial.

Estudos recentes aplicaram séries temporais, redes neurais, modelos bayesianos e algoritmos de aprendizado de máquina à malária amazônica. Esses avanços ampliam a capacidade de previsão, mas também exigem validação, transparência e interpretação epidemiológica. Um modelo com bom desempenho estatístico pode ter baixa utilidade se depender de dados indisponíveis, não informar incertezas ou gerar alertas sem possibilidade de resposta local (MONTEIRO et al., 2025).

A discussão das aplicações bioestatísticas é, portanto, relevante para pesquisadores em biologia parasitária, epidemiologia e saúde coletiva. O domínio dos métodos permite produzir análises reprodutíveis e apoiar ações territoriais, preservando a compreensão biológica e social da transmissão.

OBJETIVO

Discutir as principais aplicações da bioestatística na vigilância epidemiológica da malária na Amazônia brasileira, com ênfase em indicadores, séries temporais, regressão, análise espacial, modelos preditivos e critérios de validade.

METODOLOGIA

Trata-se de revisão narrativa e documental, de abordagem qualitativa, natureza aplicada e objetivo descritivo. Foram consultados artigos científicos e documentos técnicos sobre epidemiologia da malária, SIVEP-Malária, análise de séries temporais, modelagem espacial, regressão e aprendizado de máquina. As buscas utilizaram combinações em português e inglês dos termos malária, Amazônia brasileira, bioestatística, séries temporais, análise espacial, regressão, previsão e machine learning.

Foram priorizadas fontes publicadas entre 2020 e 2026, estudos realizados no Brasil e documentos oficiais relacionados ao monitoramento e à eliminação da malária. A síntese foi organizada segundo as etapas do processo analítico: descrição, comparação, explicação, previsão e tradução do conhecimento. Não foram efetuadas metanálise nem combinação quantitativa de estimativas, devido à heterogeneidade dos desenhos e desfechos. Como foram utilizados dados secundários publicados e documentos de acesso público, sem identificação individual, não houve necessidade de avaliação ética.

RESULTADOS E DISCUSSÃO

Estatística descritiva e indicadores epidemiológicos

A análise começa pela verificação da base: duplicidades, valores ausentes, resultados negativos, lâminas de verificação de cura, datas inconsistentes e municípios sem denominador populacional adequado. Frequências absolutas descrevem a carga de atendimento, enquanto frequências relativas permitem comparar composição por sexo,

idade, espécie parasitária e área de infecção. Mediana e intervalo interquartil são preferíveis à média quando distribuições de idade ou tempo até o tratamento apresentam assimetria.

A Incidência Parasitária Anual é um indicador central, calculado pela relação entre exames positivos e população sob risco, multiplicada por mil. Seu uso permite estratificar municípios, mas requer denominadores atualizados e compatíveis com a população realmente exposta. Em localidades de garimpo, fronteira ou intensa mobilidade, a população residente pode subestimar o contingente em risco. A proporção de casos por *P. falciparum* e infecções mistas é igualmente relevante, por indicar risco clínico e desafios de eliminação.

Indicadores de oportunidade medem o intervalo entre início dos sintomas, diagnóstico e tratamento. A análise deve apresentar proporções tratadas em períodos recomendados e comparar grupos ou territórios. Intervalos de confiança ajudam a mostrar a precisão das estimativas, especialmente quando o número de casos é pequeno. Percentuais sem numerador e denominador podem produzir interpretações equivocadas.

Séries temporais, tendência e sazonalidade

Dados mensais ou semanais formam séries temporais nas quais as observações são correlacionadas. Gráficos, médias móveis e decomposição em tendência, sazonalidade e componente irregular são etapas exploratórias. Mudanças de nível podem coincidir com intervenções, crises assistenciais, alteração ambiental ou modificação da notificação. A análise de tendência por regressão segmentada ou modelos de Prais-Winsten é útil para estimar variação percentual e identificar pontos de inflexão.

Modelos ARIMA e SARIMA representam dependências autoregressivas, diferenciação e componentes sazonais. Sua construção exige estacionariedade, inspeção das funções de autocorrelação e separação entre dados de ajuste e validação. Valiati, Rice e Villela (2024) demonstraram que a sazonalidade da transmissão por *P. vivax* varia entre áreas da bacia amazônica e é influenciada por relações climáticas não uniformes. Isso reforça que um único calendário de risco não descreve toda a região.

Séries temporais interrompidas podem avaliar políticas ou intervenções quando existe período suficiente antes e depois da implementação. A interpretação deve considerar eventos simultâneos e mudanças na cobertura diagnóstica. Laporta et al. (2022) combinaram análise espacial e temporal para avaliar o progresso brasileiro em direção às metas de eliminação, mostrando como projeções podem revelar diferenças entre tendência observada e objetivo programático.

Regressão para dados de contagem e fatores associados

A regressão de Poisson é adequada para contagens quando média e variância são próximas. Na prática, casos de malária frequentemente apresentam sobredispersão; nesse

cenário, a regressão binomial negativa produz erros padrão mais realistas. Modelos com excesso de zeros podem utilizar estruturas zero-inflated ou hurdle, distinguindo ausência estrutural de transmissão e semanas sem casos em áreas receptivas.

Variáveis ambientais podem apresentar associações não lineares e efeitos defasados. Modelos aditivos generalizados permitem curvas flexíveis para temperatura, chuva e cobertura vegetal. Termos de defasagem distribuída ajudam a representar o tempo entre alteração ambiental, desenvolvimento vetorial, infecção e notificação. Modelos mistos incorporam efeitos aleatórios para municípios ou localidades, evitando tratar observações repetidas como independentes.

Arisco et al. (2024) aplicaram modelo hierárquico bayesiano espaço-temporal com distribuição de Poisson zero-inflada para investigar desmatamento e malária entre 2003 e 2022. O estudo exemplifica como a escolha do modelo pode acomodar excesso de zeros, dependência espacial, defasagem temporal e heterogeneidade entre estados. A associação estatística, contudo, deve ser interpretada à luz de mobilidade, acesso ao diagnóstico e mecanismos ecológicos plausíveis.

Tabela 1: Métodos bioestatísticos aplicáveis à vigilância da malária.

Finalidade	Métodos	Aplicação	Cuidados principais
Descrição	Frequências, mediana, proporções, IPA	Caracterizar carga, perfil e oportunidade	Denominadores inadequados e valores ausentes
Tendência	Prais-Winsten, regressão segmentada	Estimar variação e pontos de mudança	Autocorrelação e intervenções simultâneas
Sazonalidade	Decomposição, ARIMA/SARIMA	Prever semanas ou meses de maior demanda	Séries curtas e mudanças estruturais
Associação	Poisson, binomial negativa, GAM	Quantificar fatores ambientais e sociais	Sobredispersão, não linearidade e confusão
Espaço-temporal	Moran, varredura, modelos bayesianos	Detectar aglomerados e estimar risco local	Dependência espacial e áreas sem geocódigo
Predição	Random Forest, SVR, redes neurais	Gerar alertas e planejar recursos	Sobreajuste, baixa interpretabilidade e deriva do modelo

Fonte: elaboração própria (2026).

Análise espacial e espaço-temporal

Mapas de casos absolutos podem destacar municípios populosos e ocultar áreas de alta incidência. A análise espacial deve utilizar taxas, suavização quando pertinente e avaliação da estabilidade dos denominadores. O índice de Moran identifica autocorrelação global, enquanto indicadores locais revelam aglomerados de alto ou baixo risco. Estatísticas de varredura espaço-temporal podem detectar clusters em períodos específicos, contribuindo para resposta focal.

Modelos bayesianos hierárquicos permitem combinar risco médio, estrutura de vizinhança e tendência temporal, produzindo estimativas suavizadas e intervalos de credibilidade. São úteis em municípios com poucos casos, mas podem mascarar surtos pequenos se o grau de suavização for excessivo. A unidade espacial precisa corresponder à decisão: município pode ser adequado para planejamento de insumos, enquanto localidade, território indígena ou área de mineração pode ser necessária para controle operacional.

A geocodificação incompleta limita a análise. O boletim epidemiológico nacional registra que apenas parte das localidades do SIVEP-Malária possui coordenadas, o que exige cautela na interpretação dos mapas (BRASIL, 2024a). A ausência de coordenadas não é aleatória e pode afetar principalmente áreas remotas, introduzindo viés espacial.

Modelos preditivos e aprendizado de máquina

Algoritmos de aprendizado de máquina podem capturar relações complexas e não lineares. Random Forest, Support Vector Regression, XGBoost e redes neurais recorrentes têm sido comparados com modelos tradicionais. Barboza et al. (2022) avaliaram modelos de aprendizado profundo em agrupamentos de cidades do Amazonas, demonstrando a possibilidade de adaptar previsões a padrões temporais distintos.

Monteiro et al. (2025) combinaram agrupamento K-means e seis modelos de previsão para casos semanais na Amazônia Legal. Random Forest apresentou melhor desempenho na maioria dos cenários, e o agrupamento aumentou a capacidade de representar heterogeneidade regional. O estudo também evidencia que a comparação deve usar métricas adequadas, como erro absoluto médio e raiz do erro quadrático médio, em dados de validação não utilizados no treinamento.

Bom desempenho preditivo não garante causalidade. Modelos complexos podem reproduzir padrões históricos sem explicar mecanismos de transmissão. Deve-se avaliar sobreajuste, estabilidade ao longo do tempo, calibração, sensibilidade para surtos e utilidade operacional. A validação temporal é preferível à divisão aleatória, porque simula o uso real de dados passados para prever períodos futuros. Quando possível, a validação externa em outros estados ou anos aumenta a confiança na generalização.

Qualidade dos dados, vieses e incerteza

A validade de qualquer análise depende do processo de produção dos dados. Mudanças na rede diagnóstica, falta de testes, atraso de digitação, alteração de formulários e mobilidade entre município de infecção e de notificação podem gerar quebras artificiais. Lâminas de verificação de cura devem ser identificadas para evitar contagem duplicada de episódios. Recaídas por *P. vivax* e reinfecções também exigem definição clara da unidade de análise.

Dados ausentes devem ser descritos por variável e território. A exclusão completa de registros pode enviesar resultados quando a ausência se relaciona ao desfecho ou à capacidade do serviço. Imputação múltipla pode ser considerada em análises etiológicas, desde que seus pressupostos sejam plausíveis; em vigilância rotineira, painéis de completude e busca ativa podem ser mais úteis que correções exclusivamente estatísticas.

A incerteza deve acompanhar estimativas e previsões. Intervalos de confiança ou credibilidade, distribuição dos resíduos e análise de sensibilidade permitem reconhecer limites. Valores de *p* isolados não medem relevância epidemiológica. Magnitude do efeito, precisão, plausibilidade biológica, consistência e potencial de decisão devem orientar a interpretação.

Tradução dos resultados para a vigilância e eliminação

A bioestatística tem maior impacto quando vinculada a perguntas de gestão. Previsões semanais podem orientar distribuição de testes, medicamentos, mosquiteiros e equipes. Mapas de risco podem priorizar diagnóstico ativo e investigação de focos. Análises de oportunidade identificam localidades com demora no tratamento. Estudos de série temporal avaliam se ações produziram redução sustentada.

O Plano Nacional de Eliminação da Malária estabelece vigilância como intervenção central e propõe a eliminação da transmissão autóctone até 2035 (BRASIL, 2022). Para acompanhar metas, é necessário combinar indicadores nacionais com análises submunicipais, porque reduções agregadas podem coexistir com aumento em áreas indígenas ou de garimpo. O boletim de 2024 mostrou crescimento recente e concentração diferenciada por áreas especiais, evidenciando a necessidade de estratificação (BRASIL, 2024a).

A implementação de modelos deve incluir rotinas de atualização, documentação do código, controle de versões e capacitação dos usuários. Resultados precisam ser apresentados em linguagem compreensível e associados a protocolos de ação. Alertas sem recursos ou responsabilidades definidas geram fadiga e perda de confiança. A participação de epidemiologistas, estatísticos, profissionais de laboratório, entomologistas, gestores e comunidades melhora a pertinência das análises.

CONSIDERAÇÕES FINAIS

A bioestatística amplia a capacidade da vigilância da malária ao transformar registros em medidas comparáveis, estimativas de associação, mapas de risco e previsões. Estatística descritiva e indicadores como a Incidência Parasitária Anual permanecem fundamentais, mas devem ser complementados por métodos que considerem sazonalidade, sobredispersão, excesso de zeros, dependência espacial e heterogeneidade territorial.

Séries temporais, regressões para contagem, modelos hierárquicos e aprendizado de máquina oferecem respostas distintas. A seleção não deve seguir apenas complexidade ou novidade, e sim a pergunta epidemiológica, a qualidade dos dados, os pressupostos e a capacidade de interpretação. Validação temporal, métricas transparentes e representação da incerteza são requisitos para uso responsável.

Na Amazônia brasileira, análises devem reconhecer particularidades de populações indígenas, áreas rurais, garimpos, fronteiras e localidades remotas. O fortalecimento do SIVEP-Malária, da geocodificação e da formação analítica das equipes permitirá incorporar a bioestatística à rotina decisória e apoiar ações mais oportunas rumo à eliminação da doença.

REFERÊNCIAS

ARISCO, N. J.; PETERKA, C.; DINIZ, C.; SINGER, B. H.; CASTRO, M. C. Ecological change increases malaria risk in the Brazilian Amazon. *Proceedings of the National Academy of Sciences of the United States of America*, v. 121, n. 44, e2409583121, 2024. DOI: <https://doi.org/10.1073/pnas.2409583121>.

BARBOZA, M. F. X. et al. Prediction of malaria using deep learning models: a case study on city clusters in the state of Amazonas, Brazil, from 2003 to 2018. *Revista da Sociedade Brasileira de Medicina Tropical*, v. 55, e0420-2021, 2022. DOI: <https://doi.org/10.1590/0037-8682-0420-2021>.

BRASIL. Ministério da Saúde. Elimina Malária Brasil: Plano Nacional de Eliminação da Malária. Brasília, DF: 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>. Acesso em: 13 jul. 2026.

BRASIL. Ministério da Saúde. Perfil epidemiológico da malária por áreas especiais na região amazônica, Brasil, 2013 a 2023. *Boletim Epidemiológico*, Brasília, DF, v. 55, n. 14, 2024a. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/edicoes/2024/boletim-epidemiologico-volume-55-no-14.pdf>. Acesso em: 13 jul. 2026.

BRASIL. Ministério da Saúde. Situação epidemiológica da malária. Brasília, DF: Ministério da Saúde, 2024b. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/m/malaria/situacao-epidemiologica-da-malaria>. Acesso em: 13 jul. 2026.

GARCIA, K. K. S. et al. Brazil towards malaria elimination: a time-series analysis of imported cases from 2007 to 2022. *PLOS Global Public Health*, v. 4, n. 10, e0003822, 2024. DOI: <https://doi.org/10.1371/journal.pgph.0003822>.

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, 39, 2022. DOI: <https://doi.org/10.1186/s13075-022-01000-0>.

org/10.1186/s40249-022-00945-5.

MONTEIRO, K. H. C. et al. Integrating machine learning and spatial clustering for malaria case prediction in Brazil's Legal Amazon. *BMC Infectious Diseases*, v. 25, 802, 2025. DOI: <https://doi.org/10.1186/s12879-025-11193-x>.

VALIATI, N. C. M.; RICE, B.; VILLELA, D. A. M. Disentangling the seasonality effects of malaria transmission in the Brazilian Amazon basin. *Royal Society Open Science*, v. 11, n. 7, 231764, 2024. DOI: <https://doi.org/10.1098/rsos.231764>.

WORLD HEALTH ORGANIZATION. Global technical strategy for malaria 2016-2030: 2021 update. Geneva: WHO, 2021. Disponível em: <https://www.who.int/publications/i/item/9789240031357>. Acesso em: 13 jul. 2026.

WORLD HEALTH ORGANIZATION. World malaria report 2025. Geneva: WHO, 2025. Disponível em: <https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2025>. Acesso em: 13 jul. 2026.