

FERNANDO XAVIER

**Metodologia baseada em ciência de dados para avaliação do
programa de combate à dengue**

São Paulo

2023

FERNANDO XAVIER

**Metodologia baseada em ciência de dados para avaliação do
programa de combate à dengue**

Versão Corrigida

Tese apresentada à Escola Politécnica da
Universidade de São Paulo para a obtenção do
título de Doutor em Ciências.

Área de Concentração:
Engenharia de Computação

Orientador:
Prof. Dr. Antonio Mauro Saraiva


São Paulo


2023

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, _____ de _____ de _____

Assinatura do autor: _____  Documento assinado digitalmente
FERNANDO XAVIER
Data: 29/10/2023 08:16:21-0300
Verifique em <https://validar.iti.gov.br>

Assinatura do orientador: _____  Documento assinado digitalmente
ANTONIO MAURO SARAIVA
Data: 31/10/2023 14:20:23-0300
Verifique em <https://validar.iti.gov.br>

Catálogo-na-publicação

Xavier, Fernando

Metodologia baseada em ciência de dados para avaliação do programa de combate à dengue / F. Xavier -- versão corr. -- São Paulo, 2023.

178 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1.DENGUE 2.AVALIAÇÃO EM SAÚDE 3.BIG DATA I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II.t.

Dedico este trabalho à minha mãe,
meu irmão, minha irmã e minha sobrinha,
que são as razões para eu nunca desistir.

Agradecimentos

Em primeiro lugar, agradeço a Deus, por ter me permitido chegar até aqui. Foram anos difíceis para todos, com inúmeras perdas, e chegar até aqui é uma vitória. Agradeço também a todas e todos que, com suas pesquisas, puderam desenvolver as vacinas e tratamentos contra essa doença que infelizmente vitimou tantas pessoas no mundo.

Agradeço a minha família, especialmente à minha mãe, Angela, meu irmão Luciano, minha irmã Sandra e minha sobrinha Antônia. As palavras de apoio, os ouvidos disponíveis sempre que precisei e toda a força para nunca me deixar desistir, mesmo nos piores momentos desses anos. Minha família é tudo para mim e agradeço também por toda a inspiração sobre como agir em cada momento da minha vida.

Agradeço aos meus amigos e amigas que sempre torceram por mim, pelos conselhos dados ou mesmo uma palavra amiga sempre que precisei. Também agradeço aos amigos que fiz durante essa caminhada na USP, sejam do meu departamento ou do Clube Brasileiro de Saúde Planetária. Me considero muito privilegiado por ter tantas pessoas na minha vida que gostaria de nomear aqui. Além do apoio, vocês me ensinaram muito e certamente contribuíram para que eu fosse um pesquisador melhor.

Agradeço também às pessoas participantes do Grupo de Estudos em Saúde Planetária da USP, cujas pesquisas e dicas foram muito valiosas para eu pudesse desenvolver este trabalho. Em especial, agradeço a Nancy Stoppe, e toda a equipe da Secretaria de Saúde da Prefeitura de São Paulo, por todas as informações e também por ter sido a primeira pessoa a me dizer que eu poderia usar a dengue como tema da minha pesquisa.

Também agradeço a SUCEN, em especial ao Gerson e ao Cristiano, pela parceria fundamental para que eu desenvolvesse a pesquisa. Sei que nossa parceria não acaba aqui e espero que minha pesquisa contribua para o programa de combate à dengue.

Também agradeço aos funcionários e funcionárias da USP, tão fundamentais para o funcionamento dessa instituição de excelência. Em especial, agradeço às funcionárias e funcionários do Instituto de Estudos Avançados da USP, da Escola Politécnica, assim como dos restaurantes universitários. Sentirei saudades do “bandejão”.

Esse trabalho também tem a colaboração de muitos docentes da USP, com quem tive a honra de ter reuniões e desenvolver atividades em conjunto. Em especial, agradeço as contribuições diretas ao meu trabalho da Profa. Ester Sabino, da Profa. Cristina Borba, da Profa Maria Anice Mureb Sallum, do Prof. Francisco Chiaravalloti Neto, do Prof. Fabio Cozman e da Profa. Ligia Vizeu Barrozo. Gostaria de nomear tantas outras pessoas, mas saibam que nunca esquecerei de suas contribuições.

Agradeço à CAPES, cujo apoio foi fundamental para que eu pudesse ter me dedicado ao doutorado como deve ser. Que mais estudantes possam ter acesso ao privilégio de ser bolsista dessa e de outras agências de fomento. A pandemia mostrou, mais uma vez, como o investimento em ciência é fundamental.

Por fim, não menos importante, agradeço ao meu orientador, Prof. Antonio Mauro Saraiva. Quando ouvia falar dele em Brasília, ficava curioso sobre quem era a pessoa de quem as pessoas falavam com tanta admiração. Durante esses anos, não apenas entendi isso, como admirei o Prof. Saraiva a cada dia mais. Sei que sou privilegiado em muitas coisas, e ser orientado pelo Prof. Saraiva é um desses privilégios que qualquer pessoa que convive com ele pode entender.

Além da orientação técnica, sua amizade, paciência e ouvido sempre disponíveis foram fundamentais para mim. O mundo será um lugar ainda melhor quando mais pessoas tiverem a gentileza, educação, empatia e simplicidade do Prof. Saraiva. Foi uma honra, Professor. Espero estar contigo em outros projetos.

Resumo

A dengue é uma doença que atinge milhões de pessoas anualmente no mundo. O Brasil, que tinha erradicado a doença, voltou a lidar com alto número de casos e epidemias sazonais. Apesar das ações executadas pelo poder público, não se observa uma evolução no programa de combate à dengue. A falta de compreensão das razões desse insucesso torna a dengue um problema de difícil solução. As avaliações em saúde normalmente utilizam-se de métodos que são de difícil aplicação em grande escala e nem sempre consideram fatores contextuais que podem ter influência decisiva nos resultados dos programas. Por outro lado, a era de *Big Data* tem apresentado oportunidades para avanço no conhecimento em áreas como a saúde. Nesse sentido, essa pesquisa de Doutorado teve como objetivo o desenvolvimento de uma metodologia com base em *Big Data* para avaliação do programa de combate à dengue. Nessa metodologia, buscou-se integrar fontes de dados relacionados à dengue em um repositório comum de dados, seguida da aplicação de técnicas de Ciência de Dados, como os algoritmos de aprendizado de máquina *K-Means* e *Random Forest*, além de análises estatísticas e visualização espacial. Inicialmente foi desenvolvida uma arquitetura de dados, usando conceitos de Big Data e Saúde Planetária, que foi utilizada como base para execução da avaliação do programa de combate à dengue de acordo com instrumento avaliativo desenvolvido nessa pesquisa. A solução proposta foi avaliada em um estudo de caso para os 645 municípios do Estado de São Paulo, em que foram desenvolvidos dois experimentos com utilização de técnicas de aprendizado de máquina e visualização de dados para extração de informação. Para ambos os experimentos, dados foram coletados de diversas fontes públicas e integrados em um *data lake* utilizando recursos de computação em nuvem. Além disso, foi desenvolvido um instrumento de avaliação considerando o modelo de Donabedian para avaliação de saúde que, em seguida, foi validado por especialistas de domínio. No primeiro experimento, aplicou-se o algoritmo *K-Means* para agrupar os municípios em quatro *clusters* de acordo com o contexto, como variáveis climáticas e socioeconômicas. Em seguida, foram realizadas avaliações dos indicadores dos componentes estrutura, processo e resultado do modelo Donabedian para identificar possíveis padrões. Nessas

avaliações foram comparados os valores dos indicadores em municípios do mesmo *cluster*, assim como comparações dos indicadores entre *clusters*. Para monitoramento contínuo dos indicadores, foi desenvolvido no segundo experimento um painel de *Business Intelligence* através de uma ferramenta de visualização de dados. Os resultados dos experimentos revelaram importantes informações, como a insuficiência no número de agentes comunitários de saúde nos municípios com piores resultados. Ademais, foi observado que municípios com contextos mais favoráveis à disseminação da dengue tinham deficiências em relação à estrutura de saúde local. Os resultados mostraram que o uso de técnicas de Ciência de Dados para avaliação em saúde pode agregar diversos benefícios à gestão pública. Além da análise de fatores ambientais, a metodologia proposta nesta pesquisa possibilita que as avaliações possam ser realizadas em larga escala de forma mais eficiente, agregando valor ao processo de tomada de decisão na gestão pública.

Palavras-chave: Dengue. Big Data. Donabedian. Ciência de Dados. Avaliações em Saúde.

Abstract

Dengue is a disease that affects millions of people annually worldwide. Brazil, which had eradicated the disease, has faced a high number of cases and seasonal epidemics again. Despite the public policies, there has been no improvement in the results of the dengue control program. The lack of understanding of the reasons for this failure makes dengue a problem that is difficult to solve. Health assessments typically use methods that are difficult to apply on a large scale and do not always consider contextual factors that may have a decisive influence on program results. On the other hand, the era of Big Data has presented opportunities for advances in knowledge in areas such as health. In this sense, this doctoral research aimed to develop a methodology based on Big Data for evaluating the dengue control program. In this methodology, we sought to integrate sources of data related to dengue into a common data repository, followed by the application of Data Science technique, such as K-Means and Random Forest machine learning algorithms, as well as statistical analysis and spatial visualization. Initially, a data architecture was developed, using concepts of Big Data and Planetary Health, which was used as a basis for carrying out the evaluation of the program to combat dengue according to the evaluation instrument developed in this research. The proposed solution was evaluated in a case study for the 645 municipalities in the State of São Paulo, in which two experiments were developed using machine learning techniques and data visualization for information extraction. For both experiments, data were collected from diverse public sources and integrated in a data lake using cloud computing resources. Additionally, an evaluation instrument was developed considering the Donabedian framework for health assessment and validate with domain experts. In the first experiment, the K-Means algorithm was applied to group the municipalities into four clusters according to the context, such as climate and socioeconomic variables. Then evaluations of the indicators of the structure, process, and result components of the Donabedian model were carried out to identify possible patterns. In these evaluations, the indicators values were compared in municipalities in the same cluster as well as comparisons of indicators between clusters. For continuous monitoring of indicators, a Business Intelligence panel was

developed in the second experiment using a data visualization tool. The results of the experiments revealed important information, such as the insufficient number of community health agents in the municipalities with the worst results. Furthermore, it was observed that municipalities with contexts more favorable to the spread of dengue had deficiencies in relation to the local health structure. The results showed that the use of Data Science techniques for health assessment can add several benefits to public management. In addition to the analysis of context factors, the methodology proposed in this research allows assessments to be carried out on a large scale in a more efficient way, adding value to the decision-making process in public management.

Keywords: Dengue. Big Data. Donabedian. Data Science. Health assessments.

Lista de Figuras

Figura 1 - Total de casos de dengue no Brasil entre 2001 e 2021.	3
Figura 2 - Projeção de crescimento da população urbana até 2050.	13
Figura 3 - Casos prováveis de dengue por mês do primeiro sintoma.	15
Figura 4 - Projeção de áreas favoráveis ao mosquito da dengue conforme projeções de mudanças climáticas, populacionais e socioeconômicas.	16
Figura 5 - Fluxo de informação dos casos de dengue.	39
Figura 6 - Distribuição dos municípios do estado de São Paulo quanto à série histórica de dengue, 2010 a 2019.	44
Figura 7 - Ciclo PDCA.	49
51	
Figura 8 - Componentes da intervenção.	51
Figura 9 - Tipos de análise na pesquisa avaliativa.	52
Figura 10 - Uso integrado dos tipos de avaliação.	55
Figura 11 – Componentes do modelo de avaliação de Donabedian.	56
Figura 12 - Exemplo de estruturação de um <i>Data Lake</i>	61
Figura 13 – Figura ilustrativa de um Lakehouse e suas diferenças para outros padrões arquiteturais de dados.	62
Figura 14 - Etapas da pesquisa e os métodos adotados em cada etapa.	87
Figura 15 - Fases adotadas para o desenvolvimento da solução.	90
Figura 16 – Elementos da solução proposta nesta pesquisa.	92
Figura 17 - Ciclo de Ciência de Dados usado no desenvolvimento dos experimentos.	100
Figura 18 - Taxa de mortalidade infantil para os municípios do Estado de São Paulo em 2020.	104
Figura 19 - Percentual da população atendida em saneamento entre os anos de 1999 e 2019 para o Estado de São Paulo e Brasil.	105
Figura 20 - Grau de Urbanização nos municípios do estado de São Paulo (2022). ..	106
Figura 21 - IPRS dos municípios de SP em 2018.	107
Figura 22 - Temperatura média anual para os municípios do estado de São Paulo. ..	108
Figura 23 - Municípios do estado de São Paulo infestados pelas duas espécies transmissoras do vírus da dengue.	111
Figura 24 - Taxa de incidência de casos prováveis de dengue para o estado de SP, Sudeste (sem considerar SP) e Brasil (sem considerar SP).	112
Figura 25 - Taxa de incidência para os municípios do estado de SP. Dados: média móvel de casos prováveis 2019-2021.	113
Figura 26 - Taxa de incidência para os municípios do estado de SP em função do coeficiente por porte populacional.	114
Figura 27 – Arquitetura definida para execução das avaliações.	119
Figura 28 - Avaliação do número ótimo de clusters.	130
Figura 29 - Visualização espacial dos grupos identificados para os municípios do estado de São Paulo.	132
Figura 30 - Variáveis mais importantes para definição dos grupos de acordo com o algoritmo <i>Random Forest</i>	133

Figura 31 - Variáveis mais importantes considerando a incidência como variável dependente de acordo com o algoritmo <i>Random Forest</i>	134
Figura 32 - Esquema do Data Warehouse para os indicadores de Estrutura.	147
Figura 33 - Esquema do Data Warehouse para os indicadores de Processo.....	148
Figura 34 - Esquema do Data Warehouse para os indicadores de Resultado.	149
Figura 35 - Apresentação inicial do painel de monitoramento da dengue.....	150
Figura 36 - Dados visualizados para a DRS 01.	151
Figura 37 - Indicadores do componente Resultado.....	152
Figura 38 - Visualização detalhada dos indicadores do componente Resultado.	153
Figura 39 - Indicadores do componente Estrutura.	153
Figura 40 - Visualização detalhada dos indicadores do componente Estrutura.....	154
Figura 41 - Indicadores do componente Processo.....	155
Figura 42 - Série histórica do índice de infestação em relação a indicadores do componente Processo.	156
Figura 43 - Visualização espacial do indicador Semana com casos (média de 2015 a 2019).	157
Figura 44 - Visualização espacial do índice de infestação (média de 2015 a 2019).	158

Lista de Tabelas

Tabela 1 - Objetivos e metas do PNCD.	28
Tabela 2 - Atribuições dos agentes comunitários de saúde e agentes de controle de endemias.....	35
Tabela 3 - Responsabilidades no fluxo de informação no combate à dengue.	37
Tabela 4 - Fórmulas dos índices de infestação mais utilizados.	40
Tabela 5 - Coeficiente de incidência por porte populacional.....	43
Tabela 6 – Classificação de cenários de risco de acordo com as diretrizes do programa de combate à dengue do estado de São Paulo.....	45
Tabela 7 - Temas principais de comunicação de acordo com os cenários de risco. ..	46
Tabela 8: Descrição dos tipos de avaliação.....	54
Tabela 9 – Descrição dos grupos do Índice Paulista de Responsabilidade Social...	106
Tabela 10 – Distribuição dos investimentos em saúde em SP em 2021.	109
Tabela 11 - Investimentos em saúde por habitante em SP e demais estados da Região Sudeste em 2021.	110
Tabela 12 - Fontes de dados necessárias para o desenvolvimento da pesquisa.	120
Tabela 13 - Matriz de avaliação para dengue segundo componentes do modelo de Donabedian.	124
Tabela 14 - Variáveis de contexto selecionadas para a etapa de clusterização.....	129
Tabela 15 - Distribuições dos registros entre os grupos.	131
Tabela 16 - Valores médios para os indicadores do componente Estrutura.	135
Tabela 17 - Valores médios para os indicadores do componente Processo.	136
Tabela 18 -Valores médios para os indicadores do componente Resultado.	136
Tabela 19 - Valores médios para os indicadores incluindo o contexto.....	138
Tabela 20 - Valores médios considerando o desempenho quanto ao coeficiente de incidência (em negrito, o valor mais alto e com diferença superior a 0.01)	140
Tabela 21 - Valores médios considerando o desempenho quanto ao índice de infestação (em negrito, o valor mais alto e com diferença superior a 0,01). ...	141
Tabela 22 - Níveis de agregação utilizados como filtros no painel desenvolvido...	144
Tabela 23 - Indicadores selecionados para o painel Donabedian.	145
Tabela 24 - Níveis definidos para medição do desempenho dos indicadores de Resultado.....	151
Tabela 25 - Categorias para classificação do índice de infestação.	157

Lista de Abreviaturas

ACE *Agente de Combate às Endemias*

ACS *Agente Comunitário de Saúde*

ADL *Avaliação de Densidade Larvária*

API *Application Programming Interface*

BI *Business Intelligence*

DENV *Vírus da dengue*

DW *Data Warehouse*

DL *Data Lake*

ESF *Estratégia de Saúde da Família*

FAIR *Findable, Accessible, Interoperable, and Reusable*

FIN *Ficha Individual de Notificação*

IA *Inteligência Artificial*

IBGE *Instituto Brasileiro de Geografia e Estatística*

IDH *Índice de Desenvolvimento Humano*

IPRS *Índice Paulista de Responsabilidade Social*

LGPD *Lei Geral de Proteção de Dados*

LIRAA *Levantamento Rápido de Índices para o Aedes aegypti*

NLP *Natural Language Processing*

PEAA *Plano de Erradicação do Aedes aegypti*

PIB *Produto Interno Bruto*

PNAB *Política Nacional de Atenção Básica*

PNCD *Programa Nacional de Controle de Dengue*

PSF *Programa de Saúde da Família*

SINAN *Sistema de Informação de Agravos de Notificação*

SIOPS *Sistema de Informações sobre Orçamentos Públicos em Saúde*

SP *Estado de São Paulo*

SUS *Sistema Único de Saúde*

UBS *Unidade Básica de Saúde*

Sumário

1 INTRODUÇÃO	1
1.1 Contextualização	1
1.2 Motivação.....	2
1.3 Objetivos	5
1.4 Justificativa	5
1.5 Desenvolvimento da pesquisa	6
1.6 Organização do Documento	7
2 REFERENCIAL TEÓRICO	9
2.1 Dengue	9
2.1.1 Doenças Infecciosas	9
2.1.2 O vetor.....	10
2.1.3 Transmissão.....	10
2.1.4 Doença	11
2.1.5 Fatores determinantes.....	12
2.1.5.1 Urbanização	12
2.1.5.2 Clima	14
2.1.5.3 Desmatamento	16
2.1.5.4 Outros Fatores.....	17
2.1.6 Impactos	19
2.1.6.1 Saúde	19
2.1.6.2 Econômicos	20
2.1.7 Estratégias de Combate	21
2.1.8 Vigilância.....	22
2.1.9 Controle vetorial.....	23
2.1.10 Comunicação e Educação.....	26
2.1.11 Programa Nacional de Controle da Dengue (PNCD).....	27
2.1.12 Stakeholders	32
2.1.12.1 Cidadãos	32
2.1.12.2 Profissionais de Saúde	33
2.1.12.3 Gestores	36
2.1.13 Medidas de avaliação	38
2.1.14 Normas e Diretrizes.....	40
2.1.15 Sistemas	47
2.2 Avaliação em Saúde.....	49
2.2.1 Tipos de Avaliação.....	50
2.2.2 Modelo de Avaliação de Donabedian.....	56
2.3 Big Data e Ciência de Dados	56
2.3.1 Oportunidades e Desafios.....	57
2.3.2 Ciência de Dados e Saúde Planetária	63
2.3.3 Big Data e Aprendizado de Máquina	64
2.3.4 Aplicações de ML na Saúde.....	65
2.3.5 Questões emergentes sobre dados	66
3 TRABALHOS RELACIONADOS	69
3.1 Aplicações de ML	69
3.1.1 Área da Saúde	69
3.1.2 Dengue	72

3.2 Avaliação em saúde e Avaliação em saúde – dengue	75
3.3 Características Comuns na Avaliação em Saúde	83
3.4 Limitações e Desafios	85
4 DESENVOLVIMENTO DA PESQUISA	87
4.1 Definição do problema	87
4.2 Desenvolvimento da solução: a metodologia de avaliação	90
4.2.1 Repositório de dados	93
4.2.1.1 Identificação das Fontes de Dados	93
4.2.1.2 Definição da Arquitetura do Repositório	94
4.2.1.3 Construção do Repositório	95
4.2.1.4 Ingestão dos Dados	95
4.2.1.5 Criação dos Mecanismos de Acesso	96
4.2.2 Componentes da Avaliação	97
4.2.2.1 Modelo	97
4.2.2.2 Elaboração	98
4.2.2.3 Execução	99
4.3 Avaliação da Solução	99
5 ESTUDO DE CASO	103
5.1 Área de Estudo	103
5.1.1 Características Gerais	103
5.1.2 Clima	107
5.1.3 Infraestrutura de Saúde	108
5.1.4 Investimentos	109
5.1.5 Dengue	110
5.2 Descrição dos Experimentos	114
5.2.1 Experimento 1: Avaliação Donabedian de acordo com o contexto	114
5.2.1.1 Objetivo	114
5.2.1.2 Descrição	115
5.2.1.3 Justificativa	115
5.2.1.4 Contribuições Esperadas	115
5.2.1.5 Atividades	116
5.2.2 Experimento 2: Painel BI para avaliações baseadas no modelo de Donabedian	117
5.2.2.1 Objetivo	117
5.2.2.2 Descrição	117
5.2.2.3 Justificativa	117
5.2.2.4 Contribuições Esperadas	118
5.2.2.5 Atividades	118
6 RESULTADOS	119
6.1 Fontes de Dados	120
6.2 Repositório de Dados	121
6.3 Avaliações	122
6.3.1 Matriz de Avaliação	123
6.4 Experimento 1: Avaliação Donabedian de acordo com o contexto	125
6.4.1 Descrição do experimento	125
6.4.2 Etapas	125
6.4.3 Resultados	127
6.4.3.1 Etapa 1: Seleção e preparação dos dados	127
6.4.3.2 Etapa 2: Clusterização	129
6.4.3.3 Etapa 3: Descrição das variáveis mais importantes	132
6.4.3.4 Etapa 4: Análise dos resultados usando o modelo de Donabedian	134
6.4.3.4.1 Análise intergrupos	135
6.4.3.4.2 Análise intragrupos	139
6.5 Experimento 2: Painel BI para avaliações baseadas no modelo de Donabedian	143
6.5.1 Descrição experimento	143
6.5.1.1 Objetivo	143

6.5.1.2 Etapas	143
6.5.2 Resultados	144
6.5.2.1 Etapa 1: Seleção dos indicadores	144
6.5.2.2 Etapa 2: Criação do <i>Data Warehouse</i>	146
6.5.2.3 Etapa 3: Aquisição dos dados	149
6.5.2.4 Etapa 4: Criação do Painel	149
7 CONSIDERAÇÕES FINAIS	159
REFERÊNCIAS	165
APÊNDICE A - ARTEFATOS DESENVOLVIDOS	178

1 INTRODUÇÃO

Nesse capítulo será contextualizado o problema a que está relacionada esta pesquisa, incluindo informações sobre a dengue e os diversos fatores relacionados. A seguir, serão apresentados os objetivos para o desenvolvimento da pesquisa, assim como a metodologia adotada.

1.1 Contextualização

O desenvolvimento econômico e social ao longo das décadas no Brasil tem gerado melhoria em diversos indicadores de saúde pública, como diminuição da mortalidade infantil e aumento da expectativa de vida da população. Isso está relacionado a diversos fatores, como melhorias na infraestrutura básica e desenvolvimento de intervenções para combate às doenças e promoção do bem-estar. No entanto, existe um paradoxo nesse processo de desenvolvimento, visto que muitas vezes ele é feito com base na degradação dos recursos naturais, o que pode causar prejuízos à saúde humana.

Essa questão é o aspecto chave do que se chama de Saúde Planetária, área de pesquisa que trata dos diversos fatores envolvidos na relação da saúde humana com as condições ambientais. As pesquisas nessa área mostram que a saúde não está dissociada do meio ambiente e que a degradação ambiental, embora muitas vezes feita sob o discurso de melhoria das condições de vida, pode causar impactos no bem-estar da população a médio e longo prazo (WHITMEE et al., 2015).

Um exemplo desse paradoxo entre desenvolvimento e piora no bem-estar se refere aos processos de urbanização. Enquanto a urbanização tem como objetivo a melhoria de diversos fatores locais, como saneamento básico, quando feita de maneira inadequada pode se constituir em um elemento facilitador para o surgimento de doenças. Além disso, muitas vezes o processo de urbanização não leva em consideração a preservação dos ecossistemas, tão necessários para a proteção da vida.

Com a urbanização avançando sobre áreas ocupadas por outras espécies, naturalmente há o contato dos seres humanos com diversos patógenos que podem levar ao desenvolvimento de doenças. Há indícios de que a pandemia causada pelo Sars-CoV-2, que afetou todo o planeta com milhões de infectados e óbitos, foi gerada por um processo de *spillover* (ACOSTA et al., 2020), fazendo com que o patógeno fosse transmitido de animais para os humanos. Outro exemplo de doença diretamente ligada a esse processo de degradação ambiental em nome da urbanização é a dengue. Essa doença tem como vetor o mosquito *Aedes Aegypti*, que encontra na urbanização condições favoráveis para a disseminação do vírus causador da doença.

Além da urbanização, as condições climáticas são outros fatores que podem facilitar o desenvolvimento do mosquito. Com o aquecimento global causado pela degradação ambiental, locais que antes não tinham condições adequadas para o desenvolvimento do mosquito podem se tornar mais favoráveis para a disseminação dessa e outras doenças transmitidas por mosquitos.

A piora nas condições ambientais também tem grande impacto no campo, fazendo que grandes áreas deixem de ser aptas ao cultivo de diversas espécies e, como consequência, ocasionando um movimento migratório para as áreas urbanas. Somado ao crescimento populacional, poderá haver uma degradação nas condições de vida das pessoas, além de criar mais condições favoráveis para a disseminação de doenças. O crescimento desordenado pode, por exemplo, aumentar o número de moradias com condições precárias, como falta de saneamento básico (ARAÚJO, 2018).

Dessa forma, mesmo que tenha havido melhoria em diversos índices socioeconômicos no período, quando se analisa a ocorrência de dengue, constata-se que o Brasil regrediu nesse mesmo período. Considerando-se que o processo de urbanização é um caminho sem volta e que as condições climáticas favoráveis aos mosquitos ocorrerão em mais locais e por mais tempo, o país pode enfrentar desafios ainda maiores no combate a essas doenças.

1.2 Motivação

A dengue é um problema de saúde pública mundial e estimativas indicam cerca de 390 milhões de casos por ano, sendo que, destes, 96 milhões manifestam algum tipo

de sintoma (BHATT et al., 2013). Em 1947, houve um esforço coordenado pela Organização Pan-Americana da Saúde e a Organização Mundial da Saúde para erradicação do vetor, *Aedes aegypti*, nas Américas, e o Brasil conseguiu erradicá-lo em 1955 (BRAGA; VALLE, 2007). No entanto, após a reintrodução desse mosquito em 1976 no Brasil e com mais condições favoráveis, novos casos surgiram a partir de Roraima em 1981, espalhando-se por todo o Brasil (BARBOSA et al., 2012). Dados coletados do Sistema de Informação de Agravos de Notificação (SINAN) entre 2001 e 2021 indicam um gradual aumento no total de casos de dengue no Brasil, conforme Figura 1.

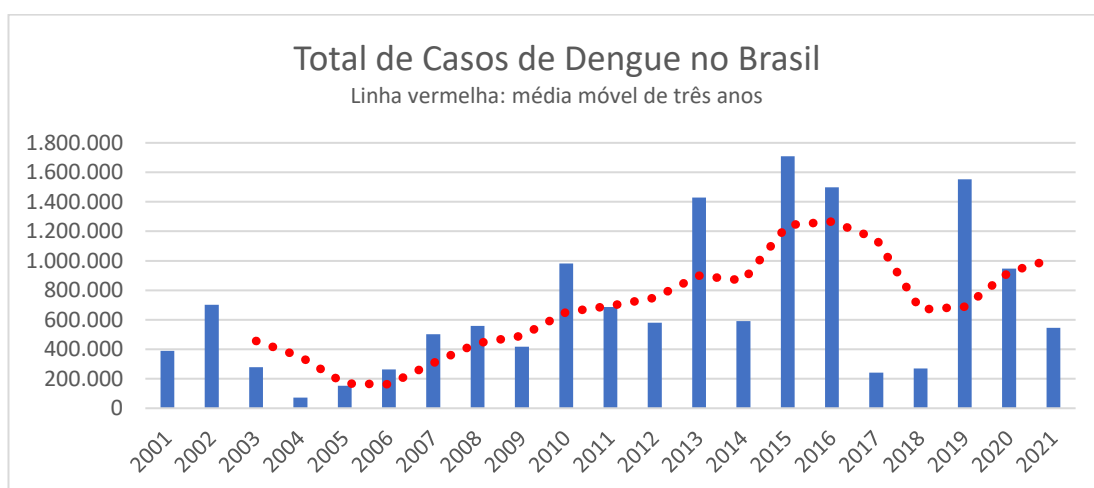


Figura 1 - Total de casos de dengue no Brasil entre 2001 e 2021.

Fonte: Dados: SINAN (SINAN, 2022).

Desde a reintrodução do vetor no país em 1976, uma série de iniciativas vêm sendo adotadas para combate da dengue culminando, em 2002, com a criação do Programa Nacional de Controle da Dengue (PNCD). No entanto, mesmo depois da adoção desse programa, a incidência da dengue no país aumentou de 389,27/100 mil habitantes em 2002, para 691,96/100 mil habitantes em 2019, aumento de cerca de 78%.

Dessa forma, a despeito da existência de um programa nacional e de grande mobilização de agentes de saúde em todo o país, não se pode afirmar que as ações de combate à dengue alcançaram êxito ainda. Podem existir algumas razões para o baixo sucesso do programa. Uma delas é que talvez as estratégias preconizadas pelo PNCD

em 2002 não sejam mais adequadas para a realidade atual, em que há muito mais mobilidade do que quando o plano foi criado. A mobilidade é um fator importante na dispersão da doença pois, uma vez que uma pessoa infectada viaja para algum lugar sem casos, basta uma picada do mosquito para que seja iniciada a transmissão em um novo local.

Outra possível razão é que as ações podem não estar sendo executadas conforme preconizadas pelo PNCD, seja por má gestão operacional dos municípios ou outras razões, como falta de pessoal. A implementação das ações do programa é realizada pelos municípios, e pode estar sujeita a diversos problemas, como (i) mau uso dos recursos públicos; (ii) mudanças constantes de equipes devido a questões políticas locais; (iii) capacitação inadequada e/ou (iv) falta de pessoal.

Fatores locais também podem influenciar a eficácia das ações, visto que as regiões têm suas particularidades, sejam elas relativas a fatores socioeconômicos ou mesmo a características ambientais. Dessa forma, uma mesma ação aplicada em locais diferentes pode não ter o mesmo resultado, não apenas por possíveis diferenças na qualidade das ações, mas talvez porque fatores locais possam contribuir para o sucesso ou fracasso de uma determinada ação. Seja por essas razões ou por outras ainda não identificadas, ainda restam diversas questões, como:

- O modelo de programa de combate à dengue está dando os resultados esperados?
- Por que, a despeito das ações adotadas, a incidência de casos de dengue continua a aumentar?
- Como aumentar a eficácia das ações realizadas?
- Quais ações tiveram maior êxito e em quais locais?
- Quais fatores podem contribuir para o sucesso das ações?

Dadas as complexas relações dos fatores relacionados à dengue, responder essas questões demanda a análise de dados provenientes de muitas áreas, o que nem sempre é possível com métodos tradicionais. No entanto, a maior disponibilidade de dados e o desenvolvimento de métodos para extração de informação no contexto de

Big Data representam oportunidades para desenvolvimento de pesquisas sobre a dengue.

1.3 Objetivos

O objeto geral deste trabalho consiste no desenvolvimento de uma metodologia com base em *Big Data* para avaliação do programa de combate à dengue. O desenvolvimento dessa pesquisa foi dividido em três partes descritas abaixo como objetivos específicos:

- Desenvolver uma arquitetura de dados Big Data-Saúde Planetária para suporte à avaliação do programa de combate à dengue;
- Elaborar um modelo de avaliação do programa de combate à dengue;
- Avaliar as ações e resultados do programa no estado de São Paulo.

1.4 Justificativa

A dengue, a despeito de muitas campanhas realizadas pelos órgãos públicos, ainda é objeto de preocupação em todo o mundo, com número crescente de pessoas infectadas anualmente. A compreensão das razões do aumento de casos, mesmo com a realização das ações preconizadas pelos programas de combate, pode trazer subsídios úteis para a formulação de políticas públicas de combate à dengue.

No entanto, a dengue é uma doença cuja ocorrência está relacionada a diversos fatores, tanto ambientais quanto socioeconômicos. Essa diversidade de fatores também está associada às políticas de combate à doença assim como suas efetivas implementações. Considerando ainda a característica tripartite do Sistema de Único de Saúde brasileiro (SUS), envolvendo os níveis federal, estadual e municipal, podem existir diferenças entre o que é preconizado pelas políticas e o que é executado de fato pelos municípios, o que pode ser investigado em pesquisas de avaliação em saúde.

Dessa forma, as respostas às perguntas descritas na subseção 1.2 sobre a efetividade do programa de combate à dengue passam por uma análise multifatorial e, por consequência, de muitas variáveis. Essas características podem representar uma oportunidade para o desenvolvimento de pesquisas que utilizem uma visão orientada

aos dados (*data-driven*) para resolução de problemas. Pesquisas com esse tipo de abordagem podem fazer uso de técnicas computacionais para extração de informação útil de grandes volumes de dados, como aprendizado de máquina.

Para a gestão pública, o uso de uma abordagem *data-driven* pode trazer importantes subsídios no planejamento e formulação de políticas públicas. Devido à complexidade de fatores envolvidos na incidência de casos de dengue (clima, uso de solo, urbanização etc.), as técnicas de análise de dados baseadas em métodos relacionados à *Big Data* podem trazer novas informações relevantes, bem como produzir respostas mais rápidas do que apenas as abordagens tradicionais para extração de informações, como análises estatísticas.

A velocidade de resposta é um fator fundamental no combate às doenças infectocontagiosas como a dengue. A exemplo do que aconteceu com a pandemia da COVID-19, o sucesso das medidas é diretamente relacionado à velocidade com que são implementadas pois, quanto mais demoradas, mais riscos há de que um surto local se transforme em uma epidemia ou pandemia. Dessa forma, produzir respostas rápidas para os gestores de saúde torna-se preponderante para o planejamento e avaliação da implementação de políticas públicas.

1.5 Desenvolvimento da pesquisa

A pesquisa foi desenvolvida utilizando a visão da Saúde Planetária, área de pesquisa que investiga a relação da saúde humana a diversos fatores, e com uma abordagem de Ciência de Dados, área de pesquisa com aplicação no contexto de *Big Data* e que conta com a participação de especialistas de domínio durante todo o desenvolvimento da pesquisa.

As atividades foram divididas em três etapas, considerando: 1) delimitação do problema, 2) desenvolvimento da solução, e 3) avaliação da solução proposta. Para a primeira etapa, foi realizado um estudo de literatura, de modo a identificar os principais aspectos relacionados à dengue, avaliações em saúde e ao tema Saúde Planetária. Além disso, nessa etapa, foram realizadas diversas reuniões com especialistas de domínio com o objetivo de identificar os principais desafios no programa de combate à dengue assim como delimitação do problema de pesquisa.

Esses especialistas foram representantes da Secretaria de Saúde da Prefeitura de São Paulo, Superintendência de Controle de Endemias da Secretaria de Saúde do Estado de São Paulo, além de pesquisadores do Grupo de Estudos em Saúde Planetária.

As reuniões com especialistas de domínio da gestão pública foram essenciais para conhecimento da prática da aplicação das ações do programa (Prefeitura de São Paulo) e sobre os principais desafios na gestão estadual do programa (Governo do Estado de São Paulo). Nas duas esferas administrativas foram identificados desafios comuns: a falta de integração dos dados e a dificuldade na extração eficiente de informações para gestão das atividades.

A interação com pessoas pesquisadoras do Grupo de Estudos em Saúde Planetária permitiu conhecer mais sobre as dimensões relacionadas à saúde, como clima, uso do solo e aspectos socioeconômicos. Pesquisadores da gestão estadual também indicaram os principais modelos para avaliação em saúde.

Na segunda etapa, foi desenvolvida a metodologia proposta para a avaliação do programa de combate à dengue. Para o desenvolvimento dessa metodologia, foram considerados as principais limitações dos métodos de avaliação existentes assim como os principais desafios para a avaliação e acompanhamento dos resultados do programa. Com uma abordagem orientada a dados, métodos de Ciência de Dados foram utilizados para a criação da metodologia de avaliação a partir dos métodos existentes na área da saúde.

Para a avaliação da metodologia proposta, foi realizado um estudo de caso para os municípios do Estado de São Paulo entre os anos de 2015 e 2019. Nesse estudo de caso foram desenvolvidos dois experimentos: (i) avaliação do programa com base em características locais, e (ii) desenvolvimento de um painel de avaliação do programa de combate à dengue. A avaliação dos resultados dos experimentos foi realizada em conjunto com os especialistas de domínio.

1.6 Organização do Documento

O próximo capítulo traz o referencial teórico relativo à saúde e aos métodos que foram adotados no desenvolvimento da pesquisa. No Capítulo 3, por sua vez, são apresentados os trabalhos relacionados ao tema desta pesquisa. No Capítulo 4 a

metodologia proposta é apresentada e são detalhados os métodos adotados para desenvolvimento e avaliação dos resultados. Em seguida, o estudo de caso realizado é apresentado no Capítulo 5, incluindo a caracterização da área de estudo assim como a descrição dos experimentos realizados. No Capítulo 6 são apresentados e discutidos os resultados dos dois experimentos desenvolvidos no estudo de caso. Por fim, no Capítulo 7 são realizadas as considerações finais deste trabalho, com discussão sobre as contribuições, limitações e trabalhos futuros.

2 REFERENCIAL TEÓRICO

Esse capítulo é dividido em três partes, iniciando com a apresentação dos conceitos relacionados à dengue, seguido pela descrição dos métodos para avaliação em saúde. Por fim, são realizadas discussões sobre *Big Data* e Ciência de Dados, com apresentação dos conceitos relacionados, aplicações na saúde e questões fundamentais para gestão dos dados no contexto de *Big Data*.

2.1 Dengue

A seguir, serão descritos os aspectos relacionados à dengue, como os fatores de transmissão, vetor, características da doenças, entre outros.

2.1.1 Doenças Infecciosas

As doenças infecciosas são causadas por agentes externos, como vírus ou bactérias. No entanto, nem toda infecção resulta em doença, podendo o indivíduo permanecer assintomático. Mesmo nesses casos, ainda pode ocorrer a transmissão para outras pessoas. A cadeia de transmissão depende essencialmente de quatro fatores: o agente infeccioso, o processo de transmissão, o hospedeiro e o ambiente (BONITA; BEAGLEHOLE; KJELLSTRÖM, 2008). O conhecimento desses fatores é fundamental para o estabelecimento das formas de combate à doença, como a definição das ações necessárias para diminuir a disseminação dos agentes infecciosos.

O processo de transmissão pode acontecer de forma direta ou indireta. Na primeira forma, o agente infeccioso pode ser transmitido imediatamente do hospedeiro para outro indivíduo, ao passo que na forma indireta a transmissão depende de um meio intermediário, como um vetor.

Exemplos de doenças infecciosas com transmissão direta incluem a AIDS e a COVID-19, cuja transmissão ocorre principalmente pelo contato entre pessoas. No caso da AIDS, a transmissão indireta também pode ocorrer, com o uso de seringas contaminadas (via parenteral) (MINISTÉRIO DA SAÚDE, 2022d). Acreditava-se, no início da pandemia da COVID-19, na possibilidade de transmissão indireta através do

contato com superfícies contaminadas (RAWLINSON; CIRIC; CLOUTMAN-GREEN, 2020) (WORLD HEALTH ORGANIZATION, 2020a). No entanto, estudos mais recentes questionam essa possibilidade, embora não seja totalmente descartada (CHENG et al., 2022; LEWIS, 2021).

Em relação às doenças cuja transmissão ocorre de forma indireta, destacam-se as arboviroses. Arboviroses são doenças causadas por vírus transmitidos por vetores artrópodes, normalmente mosquitos e carrapatos, que transmitem o vírus ao se alimentarem de sangue animal através de picadas (FIGUEIREDO, 2007). Algumas dessas arboviroses surgiram em diversas regiões do planeta em séculos recentes, como a dengue e a febre amarela, enquanto outras em regiões específicas, como a encefalite japonesa, documentada pela primeira vez em 1871 no Japão e que tem uma média de 68 mil casos de contaminação anuais (GOULD et al., 2017; WORLD HEALTH ORGANIZATION, 2019)

2.1.2 O vetor

A dengue é uma arbovirose causada pelo chamado vírus da dengue (DENV), transmitido pelos mosquitos *Aedes aegypti* e *Aedes albopictus*, que também são vetores de outras doenças como a Febre Amarela, Zika e Chikungunya (WORLD HEALTH ORGANIZATION, 2020b).

2.1.3 Transmissão

A transmissão da dengue ocorre de duas etapas: 1) de um mosquito infectado para um humano e 2) de um humano infectado para o mosquito. Ao picar um humano infectado, ocorre a replicação do vírus no intestino do mosquito e depois dissemina-se em outras partes, como as glândulas salivares, em período chamado de incubação extrínseca. Esse período, em temperaturas entre 24 e 28° C, pode levar de 8 a 12 dias, embora esse tempo possa variar de acordo com outros fatores, como genótipo do vírus e concentração viral inicial. Após esse período, o mosquito está apto a transmitir o vírus para o humano, que ocorre quando um mosquito fêmea se alimenta de sangue animal, infectando o alvo através de picadas.

Ao ser picado por um mosquito fêmea infectado, ocorre a incubação no hospedeiro, chamado de período de incubação intrínseca, e pode ocorrer de 4 a 7 dias em média (HALSTEAD, 2008). A transmissão do humano para o mosquito pode ocorrer de dois dias antes do aparecimento de sintomas até dois dias após a febre ter desaparecido. No entanto, existem evidências de que humanos assintomáticos também podem transmitir o vírus para o mosquito (DUONG et al., 2015).

2.1.4 Doença

Nos casos sintomáticos de infecção pelo vírus da dengue (DENV), os principais sintomas são: febre alta e pelo menos dois dos seguintes sintomas (WORLD HEALTH ORGANIZATION, 2020b):

- dor de cabeça severa;
- dor atrás dos olhos;
- dores musculares e articulares;
- náuseas;
- vômito;
- glândulas inchadas;
- erupção cutânea.

A dengue pode ser causada por quatro diferentes sorotipos (DENV-1, DENV-2, DENV-3 e DENV-4) e, uma vez infectado por um sorotipo, o indivíduo adquire imunidade duradoura apenas para este sorotipo, continuando suscetível a infecções dos outros três (MACIEL; SIQUEIRA JÚNIOR; MARTELLI, 2008). Infecções seguintes por outros sorotipos aumentam o risco de desenvolvimento da forma severa da doença, caso em que pode levar o indivíduo à morte. Evidências recentes indicam que infecção prévia pelo Zika vírus também constitui fator de risco para o desenvolvimento da forma severa de dengue (KATZELNICK et al., 2020).

2.1.5 Fatores determinantes

A dengue tem vários fatores relacionados à transmissão, como urbanização, clima etc. A seguir, são descritos esses fatores e incluindo dados de pesquisas realizadas sobre o impacto desses fatores na disseminação da doença.

2.1.5.1 Urbanização

A urbanização no planeta é um processo com grande aceleração nos últimos cem anos. Enquanto, há um século, apenas 20% da população mundial vivia nas cidades (NEIDERUD, 2015), uma estimativa feita pela Organização das Nações Unidas (ONU) indica que, em 2018, esse número subiu para 55,3% da população mundial (cerca de 4,22 bilhões de pessoas) e projeções para 2050 indicam que esse número subiria para 68,4%, ou 6,68 bilhões de pessoas (UNITED NATIONS, 2019).

Quando se analisa esse crescimento considerando-se o nível de desenvolvimento dos países, nota-se uma tendência de estabilização nas regiões mais desenvolvidas, ao contrário do que acontece nos países de regiões menos desenvolvidas, conforme Figura 2.

Esse crescimento da população urbana nas cidades de países menos desenvolvidos é, muitas vezes, feito de maneira desordenada, ou seja, sem planejamento adequado. Isso faz com que sejam criadas condições favoráveis para o desenvolvimento e proliferação de doenças, como aquelas causadas por mosquitos (NEIDERUD, 2015). Alguns exemplos de fatores de risco para essas doenças são:

- falta de saneamento básico;
- fornecimento precário de água;
- gestão inadequada de resíduos;
- acesso precário aos serviços públicos.

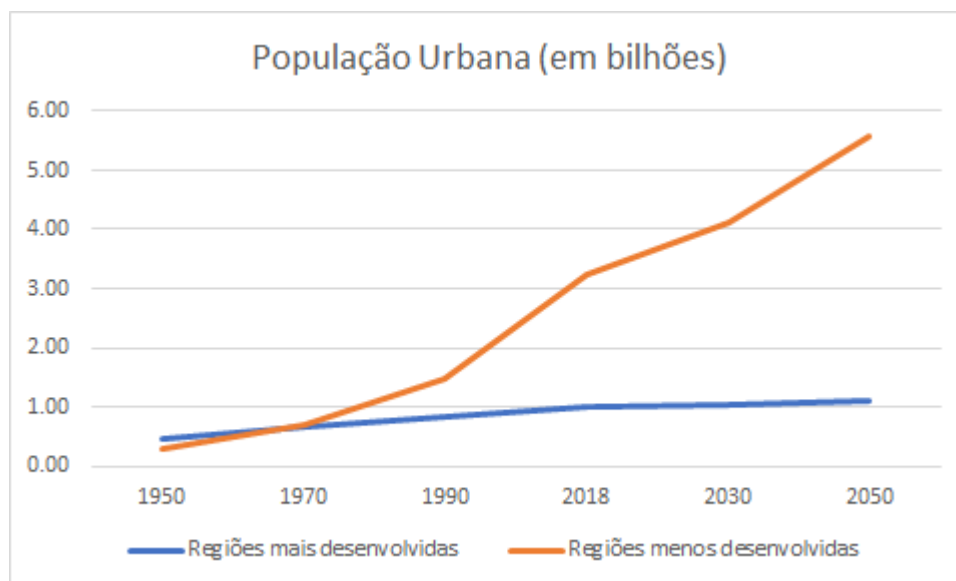


Figura 2 - Projeção de crescimento da população urbana até 2050.

Fonte: Dados: (UNITED NATIONS, 2019)

Com a ocupação desordenada das áreas urbanas, muitas pessoas vivem em moradias com condições precárias, sem serviços básicos e com alta densidade populacional. Números de 2014 indicam que cerca de 22,8% da população urbana mundial vivia em favelas, cerca de 883 milhões de pessoas no mundo (UNITED NATIONS, 2019). No Brasil, uma estimativa de 2019 indica que havia cerca de 5.127.747 domicílios em aglomerados subnormais no país (CAMPOS, 2020). Aglomerados subnormais são caracterizados pelo IBGE como “uma forma de ocupação irregular de terrenos de propriedade alheia” que são usados para habitação e normalmente sem serviços públicos essenciais (IBGE, 2020a). Geralmente, essas ocupações estão localizadas em áreas como restrições a ocupação para moradia.

Essas condições favorecem o desenvolvimento de mosquitos transmissores de doenças. Estudo realizado na China, comparando áreas urbanas e rurais, mostrou que as áreas urbanas têm diversos fatores favoráveis para a disseminação de vírus transmitidos pelo mosquito *Aedes albopictus*, como maior número de *habitats*, menor tempo de desenvolvimento larval e maior vida útil (LI et al., 2014).

Com condições tão favoráveis para o desenvolvimento do mosquito, somadas à alta densidade populacional e ao acesso precário de boa parte da população a serviços públicos, a disseminação de doenças transmitidas por esses vetores pode ser

potencializada e atingir rapidamente grande parte da população, especialmente aquelas em condições de vulnerabilidade social.

Com o aquecimento global, há o risco de intensificação do processo de migração do campo para as áreas urbanas. Isso pode ocorrer pois, com a mudança do clima, áreas antes aptas ao cultivo podem tornar-se inadequadas para a pequena agricultura, fazendo com que mais famílias migrem para as áreas urbanas em busca de emprego e melhores condições de vida. Sem políticas públicas adequadas, essa migração fará com que mais pessoas ocupem moradias com condições precárias, tornando-se mais suscetíveis para diversas doenças, como a dengue.

2.1.5.2 Clima

Diversos estudos mostram evidências das relações de fatores climáticos com o desenvolvimento e transmissão do vetor do DENV. Enquanto a temperatura pode afetar o desenvolvimento do vetor e a replicação do vírus dentro do mosquito, a variabilidade de chuvas está associada à disponibilidade de *habitats* para larvas dos mosquitos (MORIN; COMRIE; ERNST, 2013).

Estudo realizado mostrou que o tempo entre a alimentação do mosquito e a detecção do DENV-1 e DENV-4 nas glândulas salivares caiu de nove dias, em temperaturas de 26 e 28 graus, para cinco dias, em temperaturas de 30 graus (ROHANI et al., 2009). A temperatura também está relacionada a outros fatores do ciclo de vida do mosquito e alguns exemplos são: (i) ciclo reprodutivo da fêmea; (ii) sobrevivência dos mosquitos adultos; (iii) desenvolvimento dos ovos, entre outros, conforme aponta revisão realizada por Morin, Comrie e Ernst (2013) sobre os efeitos de fatores climáticos no desenvolvimento dos vetores da dengue.

A análise dos casos mensais no Brasil evidencia uma maior incidência dos casos de dengue nos períodos de maior temperatura. Dados do SINAN, de 2015 a 2019, indicam uma maior incidência entre os meses de dezembro a abril, justamente no período que compreende o início do verão brasileiro e metade de outono, quando boa parte do país ainda está com temperaturas mais altas. De maneira análoga, conforme ilustrado na Figura 3, à medida que as temperaturas caem, há uma queda dos casos de dengue no país.

Além da temperatura, a análise dos dados da Figura 3 mostra a influência de outro fator climático na ocorrência de casos de dengue: a temporada de maior ocorrência de chuvas em boa parte do país. Como o mosquito precisa de água para que os seus ovos se transformem em larva, períodos com muita precipitação podem gerar acúmulo de água em recipientes e, com as condições de temperatura ideais, são criadas as condições favoráveis para a disseminação dos mosquitos.

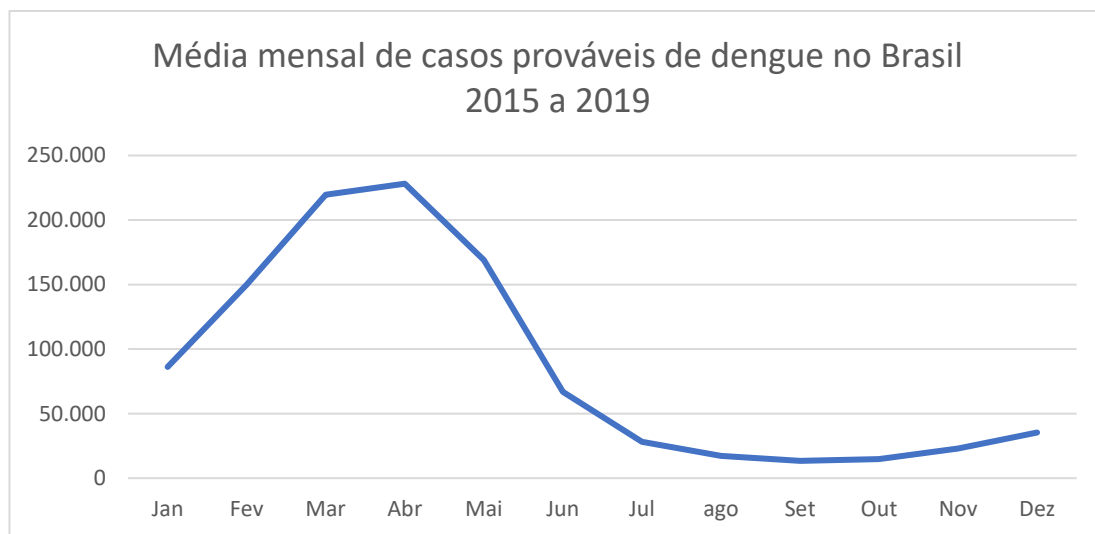


Figura 3 - Casos prováveis de dengue por mês do primeiro sintoma.

Fonte: Dados: SINAN (SINAN, 2022).

Da mesma forma que na questão da urbanização, as mudanças climáticas também podem contribuir para o aumento da incidência de dengue. Com o aquecimento global, outros períodos do ano e outros locais podem passar a ter condições favoráveis para desenvolvimento do vetor. Messina et al. (2019) realizaram modelagens para estimar o crescimento ou a diminuição de áreas favoráveis para o mosquito considerando cenários futuros, como mudanças climáticas, projeções de crescimento populacional e fatores socioeconômicos (MESSINA et al., 2019). Conforme Figura 4, para o ano de 2050, caso esses cenários se concretizem, muitas regiões do mundo terão expansão de áreas favoráveis para o mosquito. Esse mesmo estudo mostrou que os fatores relacionados ao clima (temperatura, precipitação e umidade) respondem por cerca de 86% da influência na criação de ambientes favoráveis ao mosquito.

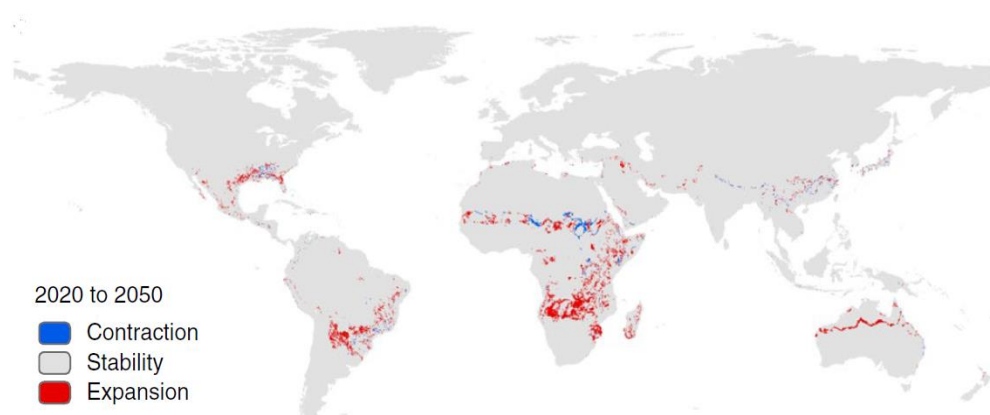


Figura 4 - Projeção de áreas favoráveis ao mosquito da dengue conforme projeções de mudanças climáticas, populacionais e socioeconômicas.

Fonte: (MESSINA et al., 2019).

2.1.5.3 Desmatamento

Por meio de uma revisão de literatura realizada identificou que, das 87 espécies de mosquito estudadas, pouco mais da metade é influenciada positivamente pelo desmatamento, incluindo um dos vetores da dengue, o *Ae. aegypti* (BURKETT-CADENA; VITTOR, 2018). Embora a mesma revisão indique que há muito a estudar sobre a relação entre desmatamento e doenças transmitidas por mosquitos, alguns fatores indiretos causados pelo desmatamento podem de fato influenciar positivamente o ciclo de vida dos insetos.

Um deles se refere às mudanças de clima causadas pela ação antropogênica no meio ambiente. As áreas verdes têm importante papel na regulação do clima em escalas regionais ou globais, como a Floresta Amazônica em relação ao clima do planeta (ELLWANGER et al., 2020). A diminuição da proteção natural causada pelo desmatamento pode aumentar a quantidade de locais com condições favoráveis para o desenvolvimento de mosquitos.

Outro fator refere-se ao uso do solo decorrente do desmatamento pois, ao destruir áreas verdes para a urbanização muitas vezes feita de forma inadequada, novas

áreas podem tornar-se favoráveis ao desenvolvimento de mosquitos urbanos e, com isso, há maior probabilidade de disseminação de doenças transmitidas por mosquitos.

Novamente ressalta-se o paradoxo ressaltado pela Saúde Planetária. Enquanto áreas verdes são destruídas em nome de um discurso desenvolvimentista para a melhoria de qualidade de vida das pessoas, essas ações podem trazer efeitos contrários, como maior disseminação de patógenos.

2.1.5.4 Outros Fatores

Dada a complexidade da doença, diversos outros fatores podem contribuir para o desenvolvimento do mosquito transmissor da dengue bem como a disseminação da doença. Ademais, esses fatores podem ganhar força em períodos esporádicos ou de maneira permanente.

Entre 2014-2015, a cidade de São Paulo passou por grande crise hídrica que fez com que a cidade enfrentasse racionamento de água. Além de um conjunto de fatores climáticos que geraram um volume de chuva muito inferior aos números históricos, essa crise também foi ocasionada pelo gerenciamento inadequado dos recursos hídricos (MARENGO et al., 2015). O racionamento de água na cidade fez com que muitas pessoas acumulassem água, muitas vezes de forma inadequada, gerando assim potenciais criadouros para larvas do *Ae. aegypti*. A cidade passou por um aumento de 163% de caso de dengue nas primeiras seis semanas de 2015, quando comparado ao mesmo período do ano anterior (WHITMEE et al., 2015).

A mobilidade é outro fator que pode contribuir para a disseminação de doenças, como a COVID-19 (ALESSANDRETTI, 2021). Estudo realizado em 2011 sobre epidemias de dengue trouxe evidências dos impactos da mobilidade humana na disseminação da doença (BARMAK et al., 2011). De acordo com resultados desse estudo, o fator da mobilidade humana não apenas influencia no tamanho da epidemia quanto também na velocidade com que ela se dissemina.

Os meses de verão são caracterizados por grande fluxo de viajantes pelo Brasil, especialmente para as cidades litorâneas. Ao ser contaminado em trânsito, um viajante pode levar o vírus para sua cidade de origem e, caso existam condições adequadas para desenvolvimento do mosquito *Ae. aegypti*, têm-se o início de uma cadeia de

transmissão em novo local. Como mostra o gráfico da Figura 3, apenas o fator mobilidade não influencia isoladamente a transmissão de casos de dengue. O mês de julho também é caracterizado, devido às férias escolares, por grande movimentação no país, mas as condições climáticas adequadas para o ciclo de vida do mosquito não são suficientes na maior parte do país.

Além do fluxo de viajantes em períodos de férias, a questão da mobilidade em relação à dengue também pode estar associada ao fluxo entre cidades por outras atividades como, por exemplo, econômicas. A influência entre as cidades ocorre em diferentes níveis hierárquicos, como metrópoles e centros locais, e pode gerar fluxo de moradores para diferentes objetivos (IBGE, 2020b). Essa movimentação intercidades também podem ser um fator facilitador para disseminação da dengue. Estudo realizado em 2018 identificou evidências que algumas cidades poderiam servir como “ponte” para transmissão em larga escala da dengue (WEN; HSU; HU, 2018).

É possível, também, que essa associação esteja mais relacionada às condições de moradia das populações mais pobres, visto que serviços como saneamento básicos e fornecimento de água são precários em muitas das localidades mais pobres. Ainda há muito a se pesquisar sobre a relação da pobreza e a dengue, segundo a revisão de literatura de Mulligan et al. (2015). De acordo com essa revisão, a análise de 12 estudos em inglês aponta evidências significativas da relação entre pobreza e dengue em apenas parte dos estudos analisados. Outro estudo, porém, relata que os riscos são altos tanto em regiões com melhores condições socioeconômicas quanto naquelas menos favorecidas (TEIXEIRA et al., 2009). Por outro lado, estudo mais recente realizado a partir de análise epidemiológica dos casos de dengue na cidade de Cambé, estado do Paraná, indica que é possível sugerir uma associação entre dengue e pobreza (BAVIA et al., 2020). Uma possível associação entre condições socioeconômicas e dengue é relacionada à taxa de mortalidade, conforme indica outra revisão de literatura, especialmente no que se refere ao acesso precário aos serviços de saúde por parte das populações mais pobres (CARABALI et al., 2015).

2.1.6 Impactos

Os impactos causados pelas doenças podem ser vistos sob diferentes perspectivas, como a saúde pública e a economia. No entanto, ressalta-se que, nas subseções a seguir, não se intenciona criar uma falsa dicotomia entre saúde e economia. Na verdade, são apresentados a seguir dados dos impactos da dengue tanto na saúde da população, que é um direito garantido pelo Artigo 196 da Constituição, quanto na economia. Os dados apresentados pelos impactos na economia causados pela dengue podem mostrar que os recursos aplicados na saúde pública devem ser vistos não como gastos, e sim como investimentos.

2.1.6.1 Saúde

Entre as doenças transmitidas por vetores artrópodes, a dengue é a segunda em número de casos e mortalidade das Américas, atrás apenas da malária (ESPINAL et al., 2019). Dados de outubro de 2019 indicam que, apenas nas Américas, houve um aumento de 13% em relação aos números de 2015, com mais de 2,7 milhões de casos e 1206 mortes (PAN AMERICAN HEALTH ORGANIZATION, 2019a). Outros dados apontam que apenas no Brasil foram mais de 2 milhões de casos e que os quatro sorotipos estão circulando pelo país (PAN AMERICAN HEALTH ORGANIZATION, 2019b).

Embora tenha havido diminuição em relação a 2019, apenas nos primeiros meses do ano de 2020 foram registrados mais de um milhão de casos no Brasil (PAN AMERICAN HEALTH ORGANIZATION, 2020). Ressalta-se o fato de que os sistemas de saúde estão com maior atenção aos casos de COVID-19 e, por isso, esses números podem estar subestimados devido à subnotificação de casos de dengue.

Além dos diversos sintomas que impactam o bem-estar dos pacientes, os riscos de óbitos devidos à dengue podem ser aumentados em situações de comorbidade com outras doenças. Um estudo identificou um aumento no risco de desenvolvimento da versão severa da dengue em pacientes com diabetes tipo 2, risco maior do que o fator idade (LEE et al., 2020). Considerando dados recentes da prevalência de diabetes do Brasil, mais de 16 milhões de pessoas têm algum tipo de diabetes e, teoricamente,

estariam em risco de desenvolver a forma mais severa da dengue (BARONE et al., 2020).

2.1.6.2 Econômicos

Além de ter impactos na saúde e bem-estar da população, a ocorrência de casos de dengue tem também impacto econômico. Embora tenha uma taxa de mortalidade menor do que outras doenças, as manifestações clínicas causadas pela dengue podem afetar a capacidade laboral das pessoas durante certo período, impactando os rendimentos necessários para subsistência de famílias, especialmente aquelas em situação de maior vulnerabilidade social.

Esse impacto pode ser minimizado com acesso adequado aos serviços de saúde, de acordo com estudo realizado por Bhalotra et al. (2019). De acordo com esse estudo, o acesso a serviços como aqueles oferecidos pela Estratégia de Saúde da Família (ESF) pode ter efeito positivo nas condições de trabalho da população atendida por esses programas (BHALOTRA et al., 2019). No Brasil, durante o período epidêmico de 2012-2013, o impacto estimado foi de US\$ 468 milhões (MACHADO, 2019). Esses impactos estão relacionados tanto aos custos adicionais ao sistema de saúde para atendimento dos casos quanto ao impacto na renda e horas de trabalho das pessoas infectadas (BHALOTRA et al., 2019).

O impacto econômico da dengue tem sido objeto de diversos estudos, analisando os custos gerados no sistema de saúde para atendimento dos casos. De acordo com dados do Sistema de Informações Hospitalares (SIH) entre 2000 e 2015, mais de 739 mil hospitalizações geraram um custo de 169 milhões de dólares, considerando apenas os atendimentos de casos de dengue e dengue severa (GODÓI et al., 2019). Considerando também que boa parte dos casos não envolve atendimento hospitalar, mas em outras instâncias do SUS, esse volume de investimento é ainda maior. Além disso, os custos indiretos, como a perda de produtividade, representam a maior parte dos custos relacionados a atendimentos aos pacientes (MARTELLI et al., 2015).

Embora o estudo realizado por Bhalotra et al (2019) tenha trazido evidências da importância de programas como a estratégia de saúde da família, a implementação

desse programa apresenta diversos desafios considerando questões como financiamento, capacitação de pessoal, entre outras, que demandam uma maior organização político-institucional (ARANTES; SHIMIZU; MERCHÁN-HAMANN, 2016). Uma política mais adequada de gestão dos recursos poderia, por exemplo, fazer com que os recursos fossem mais bem direcionados a políticas de promoção da saúde, como a ESF, gerando menos custos de atendimento nos hospitais e custos decorrentes da perda econômica pela perda de produtividade.

2.1.7 Estratégias de Combate

Ao longo do tempo, as medidas de prevenção e controle da dengue foram evoluindo da mesma forma que a dinâmica da doença. No passado, o combate à dengue era centrado em medidas de controle químico e sem grande envolvimento da população, o que se mostrou inadequado para controlar a doença (MINISTÉRIO DA SAÚDE, 2022b).

Atualmente, as medidas de prevenção e controle da dengue podem ser classificadas em quatro grupos: (1) vigilância epidemiológica e notificação dos casos; (2) controle vetorial; (3) comunicação e (4) vacinação, esta ainda em fase de desenvolvimento e testes (MACHADO, 2019). Essas ações são feitas de maneira complementar umas às outras, de forma que os resultados de uma ação possam servir de subsídio para as outras. Por exemplo, as atividades de vigilância epidemiológica podem fazer com que as ações de controle vetorial sejam intensificadas, em caso de identificação de casos em uma determinada região. De maneira análoga, ao identificar presença de vetores em um local, atividades relacionadas à comunicação podem ser realizadas de forma a educar a população para a manutenção de recipientes que podem gerar acúmulo de água.

A velocidade de integração das medidas é um fator importante para o sucesso das ações, visto que a identificação precoce dos casos pode diminuir o número de hospitalizações necessárias dos casos de dengue, bem como diminuir os riscos de morte para os pacientes infectados.

2.1.8 Vigilância

As atividades de vigilância são fundamentais para o combate a doenças como a dengue, pois provêm subsídios para o planejamento das ações de combate à dengue. Através da vigilância entomológica é possível observar e avaliar de forma contínua diversos aspectos relacionados aos vetores, como situações que favoreçam a reprodução e o impacto das intervenções (GOMES, 2002). No caso da vigilância epidemiológica, o principal objetivo é a identificação precoce dos casos da doença (TEIXEIRA; BARRETO; GUERRA, 1999). O fator tempo pode ser crucial no controle dos surtos pois, uma vez que os casos são identificados precocemente, medidas podem ser tomadas de modo a conter a disseminação dos casos.

Embora seja de notificação compulsória, a dengue, em sua forma clássica, pode ser confundida com outras doenças febris (TEIXEIRA; BARRETO; GUERRA, 1999). Com isso, nas atividades de vigilância, há a necessidade de busca ativa dos casos. Isso envolve articulação com outras áreas, como aquelas que têm maior aproximação da população, como o Programa de Saúde da Família (PSF).

Outro fator pode estar relacionado ao próprio processo de notificação, mesmo a dengue sendo um agravo de notificação compulsória. Estudo realizado em 2018 buscou identificar os fatores que podem contribuir para a subnotificação de casos através de entrevistas de 83 trabalhadores e gestores de unidades de vigilância em saúde de 17 municípios do estado de Goiás. As menções aos problemas foram classificadas em 14 categorias, com destaque para a conduta do médico ou do enfermeiro, com 23,3% das menções e para dificuldades no processo de notificação, com 13,2% (DE SOUZA MELO et al., 2018).

A qualidade dos dados informados nos sistemas de notificação pode representar uma dificuldade adicional para as atividades de vigilância. Pesquisa realizada com dados de 2014 a 2019 identificou diversos problemas de completude dos dados e concluiu que o excesso de demanda pode prejudicar o preenchimento de determinados campos (FREIRE, 2021). Outros estudos observaram problemas semelhantes nos dados notificados, como alto número de casos classificados como “inconclusivos” (DE ALMEIDA et al., 2021), e número relevante de registros duplicados (FELICIANO; CORDEIRO, 2021)

Embora existam diretrizes nacionais, a operacionalização das atividades de vigilância depende de muitos fatores, como disponibilidade e capacitação adequada de recursos humanos, recursos tecnológicos, momento epidêmico, além de normais locais.

2.1.9 Controle vetorial

Os métodos de controle vetorial podem ser divididos em três categorias: ambiental, químico e biológico. A aplicação desses métodos pode ser feita de forma combinada, com impactos nas diferentes fases do desenvolvimento dos vetores. Além disso, as condições locais e recursos disponíveis podem indicar quais são os métodos mais adequados localmente para combate ao vetor transmissor da dengue.

Pelo fato de existirem múltiplos possíveis reservatórios que servem como *a* para as larvas do mosquito, diferentes ações podem ser aplicadas de acordo com cada contexto local. Métodos ambientais podem incluir ações para modificação ou manipulação do ambiente, prevenir o desenvolvimento de larvas, assim como mudanças nas moradias ou comportamentos das pessoas, para reduzir o contato dos humanos com os vetores. Esses métodos não estão necessariamente apenas relacionados aos mosquitos, mas também a condições que podem ser favoráveis ao desenvolvimento desses além de outras doenças.

Um exemplo é a melhoria na distribuição e armazenamento de água, que não apenas está relacionada à dengue. Locais com problemas no abastecimento de água também apresentam maiores riscos para outras doenças, como cólera e leptospirose (NICHOLS; LAKE; HEAVISIDE, 2018). A má distribuição de água pode levar ao armazenamento inadequado, criando repositórios que podem servir de criadouros para larvas de mosquitos (CÁRCAMO et al., 2017). Estudos encontraram evidências de maior risco de dengue em áreas com maior frequência de escassez de água (LOWE et al., 2021).

A gestão de resíduos sólidos é outro exemplo de métodos da categoria ambiental relacionado não apenas à dengue. O acúmulo de lixo e a má gestão desses resíduos podem gerar condições favoráveis para a dengue e outras doenças, como

aquelas cujos reservatórios são roedores (DEHGHANI; OMRANI; KARRI, 2021; KRYSTOSIK et al., 2020).

Outra categoria de controle vetorial refere-se ao uso de produtos químicos para controle larval ou do mosquito na fase adulta. Para isso, as equipes envolvidas adotam estratégias que podem ser realizadas isoladamente ou combinadas, dependendo do ponto de aplicação e do estágio da epidemia. No tratamento focal, são aplicados larvicidas em depósitos com formas imaturas dos mosquitos quando a eliminação mecânica não for possível. Outra estratégia consiste no tratamento perifocal, em que o inseticida é aplicado com o uso de um aspersor manual em superfícies para atingir o mosquito que ali pousar para repouso ou desova. Essa estratégia pode ser adotada como um complemento ao tratamento focal, além de ser indicada em locais com difícil acesso. Por fim, em situações de epidemia, pode-se usar a estratégia do tratamento a ultrabaixo volume (UBV), com o uso de uma máquina pulverizadora, normalmente acoplada a veículos (MINISTÉRIO DA SAÚDE, 2022a).

Cada uma dessas estratégias tem suas vantagens e limitações. O uso de larvicidas no tratamento focal não pode ser realizado em locais que contenham peixes ou que sirvam de bebedouros de animais. No tratamento perifocal não são realizadas aplicações na parte interna em recipientes utilizados para consumo humano, como caixas d'água. O tratamento UBV tem como desvantagem a não-eliminação de cerca de 80% dos mosquitos e depende de mão de obra especializada para a aplicação. Além disso, essa estratégia pode sofrer influência das condições climáticas, como vento e chuva.

A adoção dessas abordagens de controle químico tem sua utilidade dependendo do cenário, mas podem causar outros danos, tanto para o meio ambiente como para a saúde humana. Estudo realizado em 2014 identificou efeitos tóxicos causados pela exposição aos inseticidas em diversos organismos (COLEONE, 2014). Uma pesquisa buscou analisar a nocividade para seres humanos de diversas classes de inseticidas adotados no controle vetorial e diversos sintomas foram identificados, como irritação ocular ou dérmica, efeitos em órgãos como o fígado e baço, dentro outros (GABRIEL RODRIGUES DA SILVEIRA, 2020). Além disso, essa pesquisa buscou identificar

possíveis riscos de carcinogenicidade e diversos inseticidas apresentaram algum risco quanto ao potencial carcinogênico.

Ademais, a eficácia do uso desses produtos pode ser reduzida devido a possível resistência aos inseticidas utilizados (VALLE; BELINATO; MARTINS, 2015). A diminuição da eficácia causada pela resistência aos inseticidas demanda estudos para o entendimento dos mecanismos de resistência assim como busca de formas alternativas para a utilização desse tipo de abordagem (VARGAS et al., 2022).

Dessa forma, apesar do histórico e ampla adoção no Brasil do controle químico, existem diversos questionamentos quanto à sua eficácia no controle vetorial, além dos possíveis impactos no meio ambiente e saúde humana. A experiência dos diversos programas de combate à dengue no Brasil mostrou que o controle químico não é a solução única para controlar o mosquito. No entanto, a possível baixa efetividade, limitações para uso e riscos à saúde podem demandar outros estudos para identificar as reais necessidades da adoção dessa abordagem no controle vetorial. Mais estudos são necessários para analisar se há ganhos na relação vantagens x riscos no uso de produtos químicos para o combate à dengue. Ademais, essa análise de ganhos deve considerar múltiplos aspectos, não limitando-se ao controle dos mosquitos. Além disso, deve-se considerar os impactos ao meio ambiente, pois esses podem gerar riscos à saúde humana.

Devido aos problemas e limitações do controle químico, formas alternativas de controle biológico têm sido pesquisadas. Esse tipo de controle poderia minimizar os riscos de danos ambientais do uso de inseticidas. Uma das abordagens consiste na infecção dos mosquitos com a bactéria *Wolbachia* e tem se mostrado uma forma promissora. Pesquisas indicam que os mosquitos infectados com essa bactéria têm menor capacidade de transmissão da dengue (COLLINS et al., 2022), visto que são mais resistentes à infecção pelo vírus da dengue (TANTOWIJOYO et al., 2022). Além disso, a *Wolbachia* pode manipular o ciclo reprodutivo dos mosquitos, favorecendo a continuidade dessas características para outras gerações.

Embora existam avanços no desenvolvimento de formas para controle vetorial, apenas a forma pode não ser suficiente para alcançar os resultados ideais. Existem questões relacionadas que são relacionadas à estrutura utilizada nos programas de

saúde. Um dos problemas é a falta de capacitação dos agentes envolvidos nas ações. Pesquisa realizada no município de Contagem (MG) identificou em parte dos agentes falta de conhecimento do protocolo de execução de atividades assim como em conteúdos técnicos sobre a dengue (EVANGELISTA et al., 2018). Outra pesquisa de 2018 sugere a falta de domínio técnico-científico sobre a dengue entre os agentes, com seus níveis de conhecimento aproximados ao senso comum (SOUZA et al., 2018).

2.1.10 Comunicação e Educação

A estratégia de comunicação e educação envolve a participação da população em diversas atividades, como remoção de recipientes que podem se transformar criadouros de mosquitos (MINISTÉRIO DA SAÚDE, 2022b). Para isso, a população deve receber as informações adequadas e atualizadas, como tipos de recipientes usados pelo mosquito para depósito de larvas, além de cuidados como vedação de reservatórios de água.

Para serem efetivas, as campanhas de comunicação necessitam estar adequadas a um propósito educativo e com uso adequado dos recursos de Tecnologia da Informação e Comunicação. Estudo realizado sobre as campanhas de combate à dengue, Zika e Chikungunya realizadas entre 2013 e 2017 identificou diversos problemas, como responsabilização da população pela prevenção das doenças e ênfase nas consequências das doenças, sem espaço para educação e promoção da saúde (DE ANDRADE et al., 2021).

O maior acesso à Internet através de *smartphones* representa uma oportunidade para que as atividades de comunicação alcancem mais pessoas, além de minimizar os custos de capacitação dos agentes de saúde e a população. No entanto, embora possam ser conduzidas pela esfera federal, as atividades devem considerar também as realidades locais. Nesse sentido, a gestão municipal associada aos programas de saúde da família e agentes comunitários de saúde pode ser um meio importante para adequar as informações ao contexto local.

2.1.11 Programa Nacional de Controle da Dengue (PNCD)

O combate à dengue no Brasil passou por diversas evoluções ao longo do tempo de acordo com o desenvolvimento da doença no país. Em décadas anteriores, as ações eram baseadas no combate químico ao vetor, que se mostraram insuficientes para conter a disseminação da dengue desde a reintrodução do vírus no país em 1976 (MINISTÉRIO DA SAÚDE, 2022b).

Com base nisso, as estratégias foram adequadas e, em 1996, foi criado o Plano de Erradicação do *Ae. aegypti* (PEAa), em que a principal preocupação residia nos casos de dengue hemorrágica. Esse programa, composto de nove componentes de atuação, apenas conseguiu iniciar a implantação das ações pretendidas, já que as metas não foram atendidas e houve aumento tanto de casos quanto de infestação vetorial. Embora tenha havido aumento de recursos no combate ao mosquito, as ações eram principalmente no uso de inseticidas e na eliminação de criadouros (Braga e Valle, 2007).

Estudo desenvolvido por Dias (2006) mostrou também a baixa efetividade do programa, dada a evolução espacial e temporal da dengue no território brasileiro entre 1997 e 2002, que teve sua expansão dos grandes centros para as pequenas cidades (DIAS, 2006). O estudo indica que as ações só tiveram algum êxito onde foram implementadas adequadamente e onde as condições epidemiológicas eram favoráveis.

Buscando introduzir outros elementos no combate à dengue, como participação da população, o Ministério da Saúde fez diversas avaliações dos impactos e limitações dos programas existentes. Com isso, em 2002, foi lançado o Programa Nacional de Controle da Dengue (PNCD), que buscou incorporar lições nacionais e internacionais no controle da dengue, especialmente considerando resultados dos programas desenvolvidos anteriormente (MINISTÉRIO DA SAÚDE, 2022b). São três os objetivos do PNCD e três metas relacionadas, de acordo com a Tabela 1.

Tabela 1 - Objetivos e metas do PNCD.

Objetivo	Meta
Reduzir a infestação pelo <i>Ae. aegypti</i>	Reduzir a menos de 1% a infestação predial em todos os municípios
Reduzir a incidência da dengue	Reduzir em 50% o número de casos de 2003 em relação a 2002 e, nos anos seguintes, 25% a cada ano;
Reduzir a letalidade por febre hemorrágica de dengue	Reduzir a letalidade por febre hemorrágica de dengue a menos de 1%.

Fonte: (MINISTÉRIO DA SAÚDE, 2022b).

Dadas as lições dos programas anteriores e experiências internacionais, o PNCD é fundamentado em oito aspectos essenciais:

- **a elaboração de programas permanentes**, dada a falta de evidências de que a erradicação do vetor seja possível a curto prazo;
- **o desenvolvimento de campanhas de informação e de mobilização das pessoas**, de forma a aumentar o envolvimento da população nas ações, uma vez que o *Ae. aegypti* é um mosquito essencialmente domiciliado;
- o fortalecimento da vigilância epidemiológica e entomológica para ampliar a capacidade de predição e de detecção precoce de surtos da doença;
- a melhoria da qualidade do trabalho de campo de combate ao vetor;
- **a integração das ações de controle da dengue na atenção básica**, com agentes comunitários de saúde (ACS) e Estratégia de Saúde da Família (ESF);
- **a utilização de instrumentos legais** que facilitem o trabalho do poder público na eliminação de criadouros;
- a atuação multissetorial por meio do fomento à destinação adequada de resíduos sólidos e a utilização de recipientes seguros para armazenagem de água;

- o desenvolvimento de **instrumentos mais eficazes** de **acompanhamento e supervisão** das **ações** desenvolvidas pelo Ministério da Saúde, estados e municípios.

Esses aspectos buscaram principalmente suprir as deficiências apontadas nos programas anteriores, como a não-universalização das ações e a mudança de postura do Ministério da Saúde diante da impossibilidade de erradicar o vetor, que passou a ser o controle do mesmo (BRAGA; VALLE, 2007).

A implantação do PNCD foi preconizada através de 10 componentes, adaptados a partir dos nove componentes do programa anterior (o PEAa - 1997 a 2002):

- Vigilância epidemiológica, que é dividida em quatro subcomponentes: vigilância de casos, vigilância laboratorial, vigilância em áreas de fronteira e vigilância entomológica;
- Combate ao vetor;
- Assistência aos pacientes;
- Integração com a Atenção Básica;
- Ações de saneamento ambiental;
- Ações integradas de educação em saúde, comunicação e mobilização social;
- Capacitação de recursos humanos;
- Legislação;
- Sustentação político-social;
- Acompanhamento e avaliação do PNCD.

Comparando-se com os nove componentes do programa anterior, o PNCD traz quatro novos: Assistência aos Pacientes, Integração com Atenção Básica, Sustentação político-social e Acompanhamento e avaliação, enquanto os outros foram agregados ou foram mantidos. Esses novos componentes buscaram resolver problemas anteriores

do programa, como a falta de integração entre as ações, falta de universalização e inclusão de políticas de monitoramento e melhoria contínua das ações.

Alguns estudos, no entanto, apontam para o não cumprimento das metas do PNCD. Análise dos resultados de municípios das Regiões Sudeste e Centro-Oeste, no período de 2003-2006, indica que em boa parte deles as metas não foram alcançadas (PESSANHA et al., 2009). A meta de redução de 50% dos casos de 2003 em relação a 2002 não foi alcançada em cerca de 49% dos municípios analisados e em 22% dos municípios a taxa de letalidade ficou acima da meta estabelecida.

Outro estudo, realizado para avaliar o grau de implantação do PNCD em dois municípios fronteiriços do estado do Mato Grosso do Sul, constatou que esse grau foi de 63,1% em Corumbá e 66,4% em Ponta Porã. Foram avaliados, para cada município, os graus de cada componente do programa. O componente 1 do PNCD, Vigilância Epidemiológica, foi parcialmente implantado nos dois municípios e teve como ponto crítico a falta de estrutura adequada para a implantação das ações.

Análise feita em relação aos índices de infestação predial do mosquito, em municípios da região Sul do Estado de São Paulo, identificou índices acima de 1%, número acima da meta do PNCD, bem como dificuldade de operacionalização das atividades de rotina propostas na norma técnica, com baixa cobertura e descontinuidade (DOMINGOS et al., 2013).

Em 2009, o Ministério da Saúde elaborou as diretrizes nacionais para a prevenção e controle de epidemias da dengue, um documento com detalhamento e atualizações em relação ao PNCD, criado em 2002. Esse documento contém informações sobre quatro componentes do plano (assistência, vigilância epidemiológica, controle vetorial e comunicação) para organizar as ações dos estados e municípios para atuarem de forma coordenada e uniformizada (MINISTÉRIO DA SAÚDE, 2022c).

Em relação ao componente de controle vetorial, existe uma série de ações preconizadas de acordo com o status dos municípios (infestados e não infestados). No caso dos municípios infestados, situação atual de todos os municípios do Estado de São Paulo, são previstas as seguintes ações:

- Pesquisa larvária amostral, bimestral ou quatro levantamentos rápidos de índices entomológicos (LIRAA) ao ano;
- Visita domiciliar bimestral em 100% dos imóveis;
- Pesquisa larvária nos pontos estratégicos, em ciclos quinzenais, com tratamento focal e/ou residual, com periodicidade mensal para o tratamento residual;
- Atividades de educação e comunicação;
- Articulação com órgãos municipais de limpeza urbana;
- Articulação com outros órgãos municipais governamentais e entidades não governamentais;
- Realização do bloqueio da transmissão, quando necessário.

A execução dessas ações pode estar associada à atuação das Equipes de Saúde da Família (ESF), assim como dos Agentes Comunitários de Saúde (ACS) e Agentes de Controle de Endemias (ACE). Naturalmente, como essas ações são implementadas no âmbito municipal, a execução depende da gestão de cada município e da disponibilidade de agentes para atuação, conforme preconizado nas diretrizes. Além disso, a eficácia do trabalho dos agentes depende de fatores como correta capacitação e gestão das atividades considerando-se, entre outros aspectos, a natureza integrada das ações.

Pesquisa realizada em 2016 com agentes do município de Goiânia buscou fazer uma análise de possíveis obstáculos para a execução integrada das atividades dos agentes. Nessa pesquisa, foram identificados diversos problemas, como falta de capacitação, informalidade, inexistência de protocolos etc. Outro estudo, realizado em 2007 em São José do Rio Preto, teve como objetivo a identificação de fatores de impacto na atuação dos agentes. Resultados da pesquisa apontaram dificuldades relacionadas à aceitação dos agentes pela população, dependentes da resolução de outros problemas coletivos, como falta de cuidado com as vias públicas e saneamento básico inadequado (CHIARAVALLI NETO et al., 2006). Como são também representantes do poder público, é natural que a população associe a esses

representantes as suas insatisfações em relação à atuação insuficiente de outras áreas da gestão pública.

Por conta disso, as ações de controle vetorial perdem a fundamental parceria com os moradores e o acesso às moradias para a realização das atividades de controle vetorial fica prejudicado. Na mesma pesquisa, outros problemas foram identificados como pressão para o cumprimento das metas sem as devidas condições de trabalho. Esses dois estudos, embora realizados em duas unidades da federação diferentes, ressaltam as dificuldades da implementação das ações pelos municípios por problemas de gestão local e evidenciam a necessidade da integração com ações de outras áreas da gestão pública.

2.1.12 Stakeholders

O programa de combate à dengue envolve diversos grupos de pessoas, diversos atores ou *stakeholders*, cada qual com seu nível de responsabilidade e interesses próprios que podem, muitas vezes, ser conflitantes. Assim, o sucesso das ações do programa depende da articulação adequada entre esses diversos atores. Isso se torna ainda mais importante no contexto do SUS, em que há a descentralização da gestão e das políticas de saúde. A seguir são descritos os quatro grupos envolvidos nas ações de combate à dengue: cidadãos, profissionais de saúde, pesquisadores e gestores.

2.1.12.1 Cidadãos

A população poderia ser definida como o principal grupo interessado nas ações de combate à dengue, visto que os efeitos negativos podem afetar o seu bem-estar e causar riscos à sua saúde. Embora a dengue não cause sintomas graves na maioria dos casos, o adoecimento pode prejudicar temporariamente o desempenho de suas atividades, muitas vezes responsáveis pela única forma de sustento da família.

Além disso, a participação da população é fundamental nas ações de combate à dengue. Com base nisso, o PNCD preconiza diversas ações para serem realizadas junto a população, como orientação para adequada conservação dos seus espaços para evitar a geração de criadouros para mosquitos. Estudos trazem evidência que

atividades educativas podem reduzir o número de criadouros nas residências (MADEIRA et al., 2002).

No entanto, deve-se considerar que o engajamento da população depende do acesso a informações corretas e atualizadas, visto que o vetor sofre evolução continuamente. Para isso, conforme preconizado na Política Nacional de Atenção Básica (PNAB), é de competência da União, estados e municípios o desenvolvimento de políticas de formação e capacitação dos profissionais de saúde.

2.1.12.2 Profissionais de Saúde

Há diversos grupos de profissionais de saúde que atuam em atividades relacionadas ao combate à dengue, desde a prevenção até as atividades de assistência em saúde. De acordo com o Ministério da Saúde, a Atenção Básica deve ser a principal porta de entrada e o contato preferencial dos usuários ao sistema de saúde (MINISTÉRIO DA SAÚDE, 2022e). Sendo realizada em forma de trabalho em equipe e dirigida a populações de territórios definidos pelas quais assume responsabilidade sanitária, a Atenção Básica é desenvolvida com princípios de descentralização e capilaridade, devendo estar próxima à vida das pessoas.

São compostas de equipes multiprofissionais, incluindo médicos, enfermeiros, ACS, dentre outros, cada qual com seu nível de responsabilidade e com atuação dentro dos seus limites técnicos e nos momentos adequados. Na Política Nacional de Atenção Básica (PNAB) são definidas responsabilidades de cada grupo de profissionais, embora algumas sejam comuns como, por exemplo, a manutenção atualizada dos cadastros das famílias e indivíduos nos sistemas de informação, além da realização de busca ativa e notificação compulsória de doenças e agravos, como a dengue.

A PNAB coloca a Estratégia de Saúde da Família (ESF) como estratégia prioritária para a expansão e consolidação da Atenção Básica e indica como responsabilidade de todas as esferas de governo o apoio e estímulo para a adoção dessa estratégia nos municípios. Além da ESF, a PNAB busca integrar o Programa de Agentes Comunitários de Saúde. Pesquisas indicam que essa integração pode trazer ganhos para as atividades de combate à dengue, como aumento das informações sobre

a doença, redução de criadouros, podendo otimizar recursos e evitando duplicidade de ações (CHIARAVALLOTI NETO et al., 2006).

Assim, com uma atuação mais próxima da população, os profissionais da Atenção Básica em saúde podem atuar em atividades educativas bem como identificar problemas que podem resultar em surtos de dengue, como presença de criadouros de mosquito. Além disso, também é prevista a atuação desses profissionais na busca ativa de casos de dengue, atividade que pode ser fundamental especialmente quando muitos cidadãos não procuram os serviços de saúde devido à ausência de sintomas moderados ou graves.

Existe ainda uma outra categoria de profissionais de saúde relacionada ao combate da dengue: os Agentes de Combate às Endemias (ACE). São trabalhadores que tiveram suas atividades regulamentadas em 2006 pela lei 11.350 (BRASIL, 2006) e modificada pela Lei 13.595 (BRASIL, 2018a) . Nessa lei foram descritas as formas de atuação desse grupo de profissionais, assim como do grupo dos Agentes Comunitários de Saúde (ACS). De acordo com a lei 13.595, é obrigatória a presença dos ACS na estrutura de Atenção Básica de saúde dentro do modelo de assistência multiprofissional em saúde da família e dos ACEs na estrutura de vigilância epidemiológica e ambiental. Na Tabela 2, são descritas as atribuições típicas desses profissionais.

De acordo com as atribuições definidas pelas leis e listadas na Tabela 2, observa-se que os ACSs têm uma atuação mais centrada nos indivíduos, enquanto os ACEs podem atuar nas atividades relacionados à gestão do ambiente que possa ter impacto na saúde dos moradores. No entanto, percebe-se nos dois grupos de profissionais atividades relacionadas às práticas educativas. Além disso, os ACSs têm maior interação com outras áreas de atuação, fato devido à sua inserção na estratégia da saúde da família.

De acordo com as ações preconizadas no PNCD e as atribuições dos agentes, os dois grupos podem atuar em diversas ações relacionadas ao combate à dengue, como busca ativa de infectados, medidas de controle vetorial e atividades educativas. Dessa forma, cada qual dentro do seu escopo de atuação, os agentes são elementos fundamentais para as atividades de combate à dengue.

Tabela 2 - Atribuições dos agentes comunitários de saúde e agentes de controle de endemias.

Agentes Comunitários de Saúde	Agentes de Controle de Endemias
Utilização de instrumentos para diagnóstico demográfico e sociocultural;	Desenvolvimento de ações educativas e de mobilização da comunidade relativas à prevenção e ao controle de doenças e agravos à saúde;
Detalhamento das visitas domiciliares, com coleta e registro de dados relativos às suas atribuições, para fim exclusivo de controle e planejamento das ações de saúde;	Realização de ações de prevenção e controle de doenças e agravos à saúde, em interação com o Agente Comunitário de Saúde e a equipe de Atenção Básica;
Mobilização da comunidade e o estímulo à participação nas políticas públicas voltadas para as áreas de saúde e socioeducacional;	Identificação de casos suspeitos de doenças e agravos à saúde e encaminhamento, quando indicado, para a unidade de saúde de referência, assim como comunicação do fato à autoridade sanitária responsável;
Realização de visitas domiciliares regulares e periódicas para acolhimento e acompanhamento de grupos como lactantes, idosos, crianças e adolescentes;	Divulgação de informações para a comunidade sobre sinais, sintomas, riscos e agentes transmissores de doenças e sobre medidas de prevenção individuais e coletivas;
Realização de visitas domiciliares regulares e periódicas para acolhimento e acompanhamento de situações de risco à família, grupos de risco de maior vulnerabilidade social e estado vacinal;	Realização de ações de campo para pesquisa entomológica, malacológica e coleta de reservatórios de doenças;
Acompanhamento de condicionalidades de programas sociais, em parceria com os Centros de Referência de Assistência Social (CRAS);	Cadastramento e atualização da base de imóveis para planejamento e definição de estratégias de prevenção e controle de doenças;
Outras atividades compartilhadas dentro do modelo de saúde da família, quando tem formação técnica concluída, como medidas de indicadores de pressão arterial e glicemia;	Execução de ações de prevenção e controle de doenças, com a utilização de medidas de controle químico e biológico, manejo ambiental e outras ações de manejo integrado de vetores;
Participação no planejamento e no mapeamento institucional, social e demográfico;	Execução de ações de campo em projetos que visem a avaliar novas metodologias de intervenção para prevenção e controle de doenças;
Consolidação e a análise de dados obtidos nas visitas domiciliares;	Registro das informações referentes às atividades executadas, de acordo com as normas do SUS;

Realização de ações que possibilitem o conhecimento, pela comunidade, de informações obtidas em levantamentos socioepidemiológicos realizados pela equipe de saúde;	Identificação e cadastramento de situações que interfiram no curso das doenças ou que tenham importância epidemiológica relacionada principalmente aos fatores ambientais;
Participação na elaboração, na implementação, na avaliação e na reprogramação permanente dos planos de ação para o enfrentamento de determinantes do processo saúde-doença;	Mobilização da comunidade para desenvolver medidas simples de manejo ambiental e outras formas de intervenção no ambiente para o controle de vetores;
Orientação de indivíduos e de grupos sociais quanto a fluxos, rotinas e ações desenvolvidos no âmbito da atenção básica em saúde;	
Planejamento, o desenvolvimento e a avaliação de ações em saúde;	
Estímulo à participação da população no planejamento, no acompanhamento e na avaliação de ações locais em saúde.	

Fonte: Adaptado de (MINISTÉRIO DA SAÚDE, 2022c)

2.1.12.3 Gestores

De acordo com a Lei nº 8.080, o SUS é responsabilidade dos três entes federativos: União, estados e Distrito Federal, e municípios (BRASIL, 1990). A lei enfatiza a descentralização através da municipalização dos serviços e ações da saúde, devendo haver para os municípios a transferência de poder, competências e recursos.

A Portaria nº 1.378 estabelece as competências de cada ente federativo, na qual a União é responsável por diversas atividades, como normalização técnica, coordenação dos sistemas nacionais de informação de interesse da vigilância em saúde, entre outras (BRASIL, 2013). Aos Estados competem ações complementares às de responsabilidade da União, além de outras como provimento de insumos estratégicos e atividades de coordenação junto aos outros entes. Os municípios também podem complementar ações dos outros entes dentro dos seus limites, além da execução e monitoramento das ações de vigilância em seu território. Algumas responsabilidades (Tabela 3) são comuns aos três entes federativos, como promoção e fomento à

participação social nas ações de vigilância e participação no financiamento das ações de vigilância em saúde.

Tabela 3 - Responsabilidades no fluxo de informação no combate à dengue.

Município	Estado
Receber as fichas de todos os casos suspeitos notificados pelas unidades de saúde	Verificar se os dados dos municípios estão sendo atualizados semanalmente
Incluir todos os casos suspeitos no SINAN	Acompanhar a curva dos casos, a tendência e o perfil da doença, em todos os municípios do estado
Investigar todos os casos notificados	Divulgar as diretrizes técnicas de orientação aos municípios sobre notificação e investigação de casos, investigação de óbitos, coleta de amostras para sorologia e isolamento viral
Acompanhar a curva dos casos, a tendência e o perfil da doença	Estabelecer com o laboratório central a rotina para a coleta de amostras para o monitoramento da circulação viral
Comunicar imediatamente a vigilância entomológica para providências de controle vetorial	Realizar o controle de qualidade dos exames sorológicos realizados por laboratórios descentralizados
Preencher a ficha de investigação de dengue, encerrar o caso oportunamente	Realizar, por intermédio do laboratório central, exames sorológicos, de acordo com as normas definidas, quando não for possível ou indicado a realização dos testes de forma descentralizada
Investigar todos os óbitos suspeitos de dengue	Apoiar a investigação de casos graves e óbitos
Avaliar a consistência dos casos registrados no SINAN	Avaliar a consistência dos casos registrados no SINAN
Consolidar os dados municipais e produzir boletins mensais disponibilizando informações para as unidades de saúde e o público	Consolidar os dados do estado e produzir boletins mensais disponibilizando informações para os municípios e o público em geral
Enviar os dados à Secretaria Estadual de Saúde	Enviar os dados ao Ministério da Saúde
Capacitar em vigilância epidemiológica as equipes das unidades de saúde	Capacitar as equipes de vigilância epidemiológica municipal

	Prestar assessoria técnica às Secretarias Municipais de Saúde
--	---

Fonte: Adaptado de (MINISTÉRIO DA SAÚDE, 2022c)

Em relação à dengue, existem diretrizes específicas para os entes federativos. Para planejamento e tomada de ações nos períodos epidêmicos e não-epidêmicos, são necessárias diversas atividades relacionadas ao fluxo de informação.

Cabe ao município, por exemplo, informar no SINAN os dados de notificação de casos de dengue e transmitir para a vigilância estadual. Esta, por sua vez, é responsável por transmitir essas informações para o Ministério da Saúde. A Tabela 3 lista as atribuições dos Estados e municípios para o fluxo de informação nos períodos não-epidêmicos.

Em relação ao controle vetorial, cabe aos municípios a operacionalização das ações, incluindo atividades de gestão, planejamento e execução. Como atividades de gestão, inclui-se manutenção dos equipamentos utilizados, gerenciamento da força de trabalho, dentre outras. Ainda cabe aos municípios a execução de diversas atividades, como levantamento de indicadores entomológicos, integração das equipes de saúde da família, coleta e envio para os laboratórios de amostras de sangue. Em relação aos Estados, estes devem prestar assessoria técnica aos municípios, além da supervisão, avaliação e monitoramento das ações realizadas. Além disso, podem apoiar os municípios com pessoal, insumos e equipamentos em emergências. Para a União, são definidas algumas atividades, como normatizar tecnicamente as ações de vigilância e controle da dengue, consolidar os dados oriundos dos Estados além de prestar assessoria técnica.

2.1.13 Medidas de avaliação

Existem diversas medidas que podem ser utilizadas para avaliar a eficácia das ações nos componentes do PNCD. De maneira geral, relacionada a todos os componentes, usa-se o número de casos de dengue como medida básica. Outras medidas podem ser derivadas do número de casos, como o total por faixa etária ou por estrato socioeconômico. Como em outras doenças, essas avaliações permitem

identificar aspectos relacionados a determinados grupos e servir como insumo para definição de políticas públicas.

De acordo com o fluxo da Figura 5, os dados de casos são gerados no âmbito municipal, através do preenchimento de formulário padronizado no sistema SINAN WEB, e posteriormente consolidados nas esferas estadual e federal. Como em outras estatísticas de saúde, também há o problema da subnotificação em relação aos casos de dengue. Como a alimentação dos dados é iniciada pela procura do indivíduo aos serviços de saúde, pode existir uma série de casos não notificados daqueles que não procuram esses serviços por motivos diversos, como ausência de sintomas moderados ou graves.

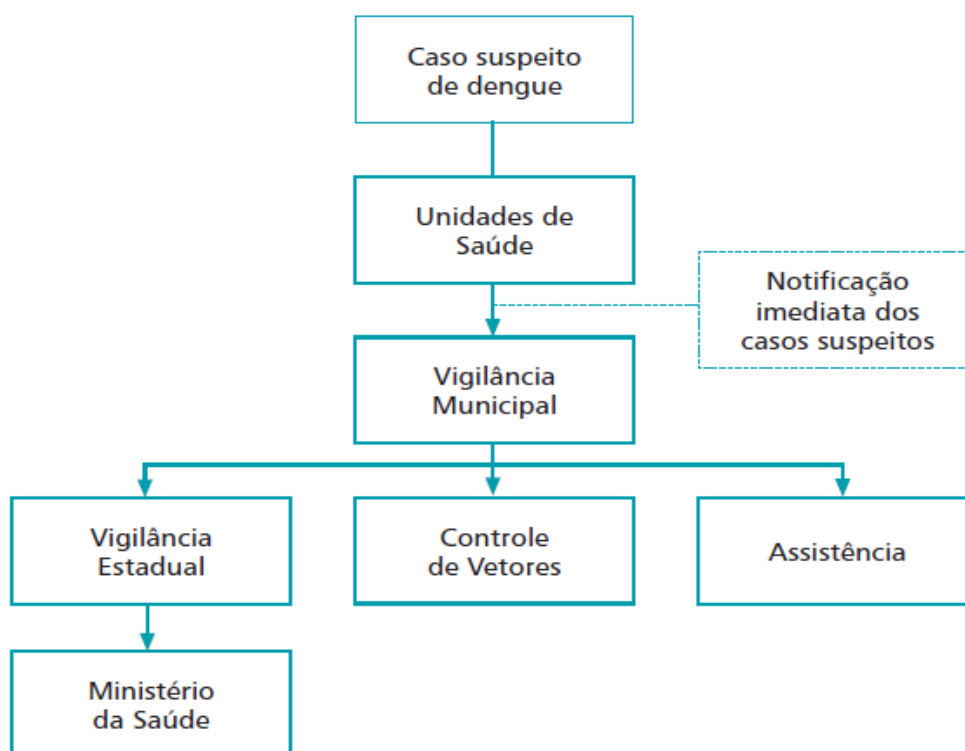


Figura 5 - Fluxo de informação dos casos de dengue.

Fonte: (MINISTÉRIO DA SAÚDE, 2022b).

Outra medida refere-se ao componente do PNCD Controle Vetorial, chamada de índice de infestação. Existem algumas formas de cálculo desse índice, sendo três as

mais utilizadas: Índice Predial, Índice de Breteau e Índice por Tipo de Recipiente, com as respectivas fórmulas detalhadas na Tabela 4.

Tabela 4 - Fórmulas dos índices de infestação mais utilizados.

Medida	Fórmula
Índice Predial	$IP = \text{Imóveis Positivos} / \text{Imóveis Pesquisados} \times 100$
Índice de Breteau	$IB = \text{Recipientes Positivos} / \text{Imóveis Pesquisados} \times 100$
Índice por Tipo de Recipiente	$ITR = \text{Recipientes Positivos} / \text{Total de recipientes positivos} \times 100$

Essas medidas podem ser usadas isoladamente ou de maneira conjunta como, por exemplo, uma avaliação mais geral como o índice predial e outra mais específica por tipo de recipiente. O uso dessas medidas pode direcionar as ações de controle vetorial a serem executadas, como alguma mais relacionada a um tipo específico de recipiente que esteja com um índice maior do que os demais. Da mesma forma, a análise pode ser feita por tipo de imóvel, como aqueles chamados de “imóveis especiais” ou os pontos estratégicos, que podem contribuir para maior disseminação dos casos de dengue. Imóveis com grande circulação de pessoas (imóveis especiais) podem acelerar a disseminação da dengue (BARBOSA et al., 2019) e uma análise da infestação por tipo de imóvel pode gerar informações para priorizar a execução das ações de controle vetorial.

2.1.14 Normas e Diretrizes

A referência nacional para a definição das ações de combate à dengue é o documento de Diretrizes Nacionais para a Prevenção e Controle de Epidemias de Dengue (MINISTÉRIO DA SAÚDE, 2022c), em que constam as normas para quatro componentes do PNCD: Assistência, Vigilância Epidemiológica, Controle Vetorial, e Comunicação e Mobilização. Esse documento indica diversos critérios para a classificação de municípios, pacientes e situação epidêmica. As classificações, seja quanto à gravidade da doença como a situação de infestação no local, são base para a definição e seleção das ações prioritárias em cada componente do PNCD. As ações

realizadas pelas secretarias de saúde visam, principalmente, controlar o surto da doença e evitar a evolução dos casos de dengue para as formas graves da doença.

Em relação à infestação de mosquitos, os municípios são classificados em dois estratos sendo, 1) municípios infestados, com disseminação do vetor nos domicílios, e 2) municípios não infestados, onde não há a presença disseminada do vetor nos domicílios. Os municípios do primeiro estrato podem ser reclassificados no segundo após 12 meses consecutivos sem a presença do vetor.

Em relação à classificação em relação à gravidade da doença, o objetivo é oferecer a melhor assistência ao paciente e, conseqüentemente, evitar óbitos. Nesse sentido, os casos são classificados de acordo com os sintomas apresentados, além da verificação de possíveis sinais de alarme que possam indicar a forma grave da doença. Os pacientes são classificados em quatro grupos (de A, mais leve, a D, mais grave) e, para cada grupo, são definidos os procedimentos de atendimento médico, como exames necessários e tratamentos indicados. Além disso, é definido o tipo de atenção à saúde que deve ser indicado ao paciente de acordo com o grupo a que pertence. Pacientes do Grupo A, por exemplo, tem as unidades de atenção primária à saúde como unidades de referência para o seu atendimento. Caso exista a evolução da doença para estágios mais graves, as diretrizes indicam a necessidade de transferência para as unidades de atenção secundária (pacientes do Grupo B), ou atenção terciária (pacientes do Grupo C, unidades com leitos internação e pacientes do Grupo D, unidades com UTI).

No fluxo de atendimento ao paciente, os dados devem ser registrados desde o início do atendimento, mesmo quando o caso ainda é suspeito de dengue, até o fechamento do caso. Para isso deve ser preenchida a ficha de notificação individual dos casos, documento que deverá ser atualizado de acordo com cada alteração no quadro como, por exemplo, a confirmação ou não da dengue.

As diretrizes ainda indicam a integração dos serviços de assistência com os demais serviços, como o controle vetorial. Uma vez que o caso suspeito é identificado na atenção primária, deve-se providenciar a visita domiciliar dos agentes comunitários de saúde para acompanhamento dos pacientes e familiares. Outras responsabilidades das unidades de atenção primária incluem identificação e eliminação de criadouros e

promoção de ações de educação em saúde e mobilização em saúde. Dessa forma, a integração das atividades das unidades da atenção primária com as atividades dos agentes é primordial para controle vetorial, assim como a busca ativa de casos e acompanhamento dos pacientes.

Além da organização das unidades de saúde, são recomendadas ações de capacitação dos profissionais de saúde, para assistência ao paciente, e agentes comunitários de saúde, para realizações de ações de prevenção e controle. Uma vez que os casos são identificados, a realização de ações oportunas pode contribuir para a redução da letalidade da doença assim como controle da disseminação. Nesse sentido, o fluxo de informação entre as unidades de saúde e os órgãos de gestão é fundamental para o processo de tomada à decisão no combate à dengue.

Outra classificação também é definida quanto ao tipo de imóvel. Além dos domicílios, existem outros tipos de imóveis que demandam atenção das equipes de combate à dengue. Essa classificação é utilizada para definição da forma e frequência das ações que serão realizadas nesses imóveis. Os chamados Pontos Estratégicos são os imóveis com potencial de geração de criadouros para os mosquitos, como borracharias, depósitos de materiais de construção, ferros-velhos etc. Nesses locais, as atividades de vigilância e controle vetorial devem ser feitas em ciclos quinzenais, além da aplicação de produtos mensalmente ou sempre que forem identificados focos.

Imóveis Especiais são aqueles onde há grande circulação de pessoas, como prédios públicos, escolas, igrejas etc., aumentando a probabilidade da disseminação de arboviroses. Dadas as suas características, esses imóveis podem ter ações específicas para identificação e controle/eliminação dos criadouros, assim como requerem integração entre seus responsáveis e os serviços de combate à dengue.

Em relação ao período, os municípios podem ser classificados em epidêmicos e não-epidêmicos. Um município pode passar pelas duas classificações durante o ano, em que a transição para o período não-epidêmico pode ocorrer devido a diversos fatores, como sucesso nas ações de controle, variações climáticas, como a diminuição de temperatura nos meses mais frios do ano e, ainda, a ausência de indivíduos suscetíveis. O período epidêmico pode ocorrer nas condições opostas, ou seja, ausência ou baixa eficácia das medidas de controle, início do período mais quente do ano e

presença de suscetíveis, que pode ocorrer pela entrada de um novo sorotipo do vírus no local.

A avaliação da situação epidêmica no local, para a devida tomada de decisão, deve ser feita continuamente através do monitoramento da curva de casos suspeitos. Pelo fato de a dengue ser uma doença com ocorrência sazonal, a análise de casos suspeitos é realizada com a comparação ante as semanas anteriores e em relação ao mesmo período dos anos anteriores. Para isso, são adotados normalmente dois tipos de mecanismos: histograma e diagrama de controle.

Segundo diretrizes do Estado de São Paulo, a aplicação de cada instrumento depende da situação do município em relação à série histórica de casos (SECRETARIA DE SAÚDE DE SÃO PAULO, 2017). Os municípios **sem série histórica** são aqueles que não atingiram 20% da incidência acumulada de casos confirmados de dengue por porte populacional em pelo menos 5 dos últimos 10 anos. Nos casos dos municípios classificados **com série histórica**, a incidência acumulada anual atingiu pelo menos 20% da incidência por porte populacional em no mínimo 5 dos últimos 10 anos. A avaliação por porte populacional é adotada devido ao fato de que mesmo um pequeno número de casos pode gerar uma incidência muito alta, em municípios com populações menores. Dessa forma, o Estado de São Paulo adota um coeficiente de incidência por número de habitantes, conforme mostrado na Tabela 5.

Tabela 5 - Coeficiente de incidência por porte populacional.

Coeficiente de incidência / 100 mil habitantes	População (número de habitantes)
600 casos	<=9.999
300 casos	Entre 10.000 e 99.999
150 casos	Entre 100.000 e 249.999
100 casos	Entre 250.000 e 499.999
80 casos	>=500.000

Fonte: SES-SP (SECRETARIA DE SAÚDE DE SÃO PAULO, 2017)

De acordo com dados entre 2010 e 2019 para o Estado de São Paulo, cerca de 57% dos municípios tinham série histórica de casos, enquanto os outros 43% (277 municípios) tiveram pelo menos 5 anos com menos de 20% da incidência durante o período analisado. Analisando-se o mapa na Figura 6, nota-se um certo padrão

espacial na distribuição dos grupos no estado. Isso pode estar relacionado a fatores climáticos assim como socioeconômicos.

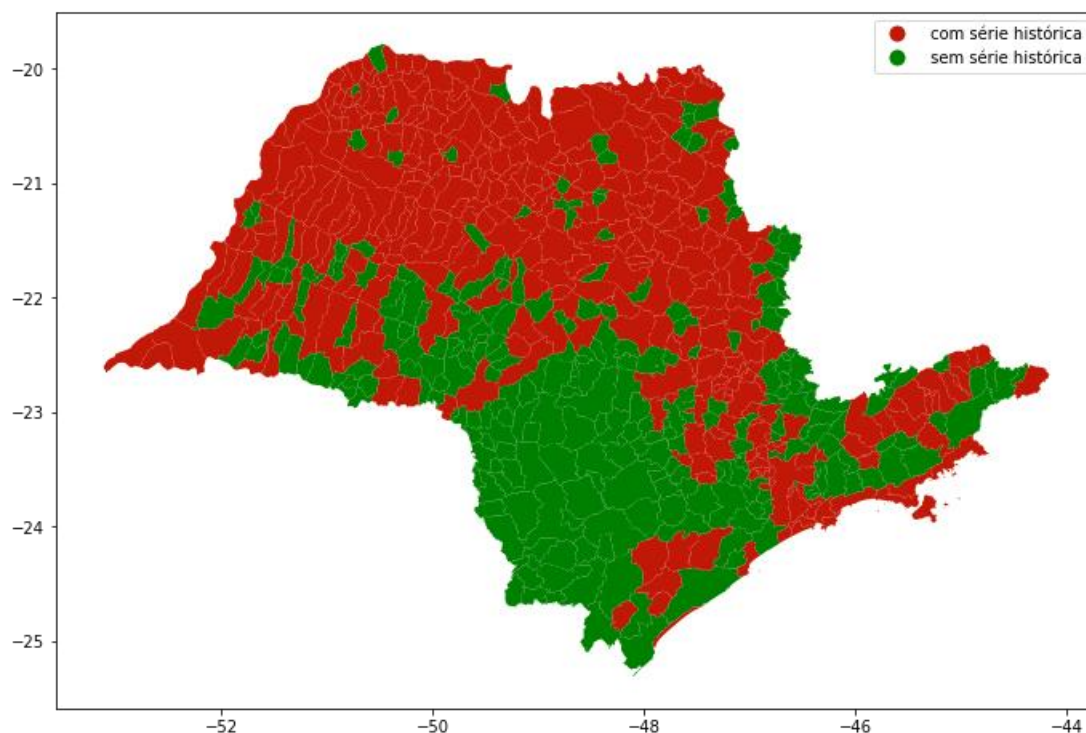


Figura 6 - Distribuição dos municípios do estado de São Paulo quanto à série histórica de dengue, 2010 a 2019.

Fonte: Dados: SINAN (SINAN, 2022).

Para cada grupo de municípios, são preconizados mecanismos específicos de acompanhamento da curva de casos, de acordo com as diretrizes do Estado de São Paulo. Para os municípios sem série histórica, a análise é feita através de histograma das quatro semanas anteriores à semana de cálculo. No momento em que a incidência atinge 20% do coeficiente por porte populacional e a tendência de alta permanece por quatro semanas consecutivas, deve ser avaliado se a coleta para confirmação será suspensa, visto sua menor utilidade em períodos epidêmicos (SECRETARIA DE SAÚDE DE SÃO PAULO, 2017).

No caso dos municípios com série histórica, a avaliação da tendência é realizada através do diagrama de controle, que considera a evolução em relação à série histórica. Para a geração do diagrama de controle, os seguintes passos são definidos:

- Seleção da incidência de casos prováveis distribuídos por semana epidemiológica dos últimos 10 anos;
- Exclusão dos anos epidêmicos e daqueles com incidência muito baixa;
- Cálculo da mediana por semana epidemiológica;
- Cálculo dos limites inferior (percentil 25) e limite superior (percentil 75).

Indica-se que está ocorrendo uma epidemia quando os valores da semana corrente ultrapassam o limite superior, enquanto os valores observados entre os dois limites caracterizam o nível endêmico, ou seja, dentro da variação esperada. Quando esse limite é ultrapassado, a tendência deve ser observada pelas quatro semanas seguintes e, em caso positivo, avalia-se a suspensão dos testes de confirmação.

O Estado de São Paulo ainda define uma classificação quanto a cenários de risco para ser utilizada na vigilância permanente. Essa classificação serve como um indicador para definição das ações a serem tomadas pelas secretarias. São, conforme descrito na Tabela 6, quatro cenários:

Tabela 6 – Classificação de cenários de risco de acordo com as diretrizes do programa de combate à dengue do estado de São Paulo.

Cenário de Risco	Parâmetro
Silencioso	Município sem notificação de casos suspeitos ou com incidência abaixo do limite inferior esperado pelo diagrama de controle
Risco Inicial	Município com incidência inferior a 20% do limite estabelecido para seu porte populacional (histograma) ou com incidência entre o limite inferior e a mediana de casos esperados pelo diagrama de controle
Risco Moderado	Município com incidência maior ou igual a 20% do limite estabelecido para seu porte populacional (histograma) ou com incidência entre a mediana e o limite superior de casos esperados pelo diagrama de controle
Alto Risco	Município que atingiu o limite estabelecido para seu porte populacional (histograma) ou com incidência acima do limite superior de casos esperados pelo diagrama de controle

Fonte: Adaptado de (SECRETARIA DE SAÚDE DE SÃO PAULO, 2017)

Para cada cenário de risco são definidas ações específicas no componente de Comunicação e Mobilização. De acordo com a Tabela 7, as ações para os cenários de menor risco têm um foco maior no controle vetorial, enquanto nos de maior risco o objetivo principal é evitar o agravamento dos casos e óbitos.

Tabela 7 - Temas principais de comunicação de acordo com os cenários de risco.

Silencioso e Risco Inicial	Risco Moderado e Alto Risco
a eliminação dos criadouros do vetor.	sinais e sintomas de complicação.
a biologia e os hábitos do vetor.	perigos da automedicação.
os locais de concentração do agente transmissor.	orientação quanto ao atendimento médico na unidade de saúde mais próxima ou unidades de referência logo nos primeiros sintomas.
os principais sintomas das doenças; recomendações sobre medidas proteção individual, especialmente para as gestantes.	importância da notificação de suspeitos; esclarecimentos sobre medidas de autocuidado, especialmente sobre a hidratação oral.
recomendações para que a população recorra aos serviços de atenção primária à saúde, no caso de surgimento de sinais e sintomas.	esclarecimentos sobre medidas proteção individual, especialmente para as gestantes.
	reforço às ações de eliminação de criadouros do vetor especialmente quanto à remoção de depósitos.

Fonte: Adaptado de (SECRETARIA DE SAÚDE DE SÃO PAULO, 2017)

O monitoramento contínuo da situação epidemiológica fornece às secretarias de saúde importantes indicadores para apoio à tomada de decisão. Como em outras doenças infectocontagiosas, o sucesso no controle da disseminação depende da velocidade com que as ações são tomadas. Para isso, os dados devem ser facilmente integrados e o fluxo de dados entre as unidades de saúde deve ser o mais rápido possível. Devido ao fato de a dengue ser uma doença com poucos casos graves, a busca ativa de casos deve estar integrada com as atividades da atenção básica em saúde e dos agentes comunitários. O correto diagnóstico do cenário de risco é fundamental para determinação das medidas adequadas para o momento.

2.1.15 Sistemas

O principal sistema de informação relacionado ao combate à dengue é o Sistema de Informação de Agravos de Notificação (SINAN) e é alimentado principalmente pela notificação de casos de doenças de notificação compulsória (SINAN, 2022). A alimentação dos dados ocorre de forma descentralizada, com os municípios sendo responsáveis pelo preenchimento e transmissão dos dados para as respectivas secretarias estaduais de saúde e, estas, são responsáveis pela consolidação e transmissão para o Ministério da Saúde.

Pode ser utilizado nas unidades de saúde e, na ausência de recursos computacionais nessas unidades, os dados podem ser preenchidos nas secretarias municipais ou, na ausência de recursos computacionais nos municípios, nas administrações regionais de saúde. O documento básico de entrada é a Ficha Individual de Notificação (FIN), que deve ter preenchimento para cada paciente com suspeita de ocorrência de doença de notificação compulsória. É facultado aos estados e municípios o uso do SINAN para inclusão e dados de outros agravos de interesse local. Os dados preenchidos nos municípios devem ser enviados semanalmente para as secretarias estaduais, que são responsáveis por encaminhamento quinzenal dos dados para o Ministério da Saúde.

Na ausência de suspeitas de doenças, as unidades necessitam preencher um documento de notificação negativa. A falta de alimentação dos dados pelos municípios por dois meses consecutivos pode gerar sanções, como suspensão no repasse de recursos.

Embora compreenda agravos de notificação compulsória, o SINAN está sujeito a problemas de subnotificação. Uma revisão de literatura analisou estudos relacionados ao tema (DE SOUZA MELO et al., 2018) e, dentre os principais fatores associados à subnotificação, destacam-se:

- Notificação em mais de um sistema;
- Falta de padronização do código internacional de doenças (CID) entre sistemas;
- Não integração dos sistemas;

- Falta de experiência ou conhecimento suficiente dos profissionais;
- Prioridades diárias dos profissionais de saúde;
- Menor integração da vigilância com a rede privada;
- Volume em epidemias;
- Perda de seguimento dos casos.

Pelo fato de a administração ser descentralizada, é natural que os municípios implementem suas próprias formas de controle que possam ser mais adequadas às suas realidades, sendo a utilização do SINAN uma forma de cumprir uma obrigação e não estar sujeitos a sanções. Embora seja desafiador o desenvolvimento de um sistema que atenda às especificidades de um país tão diverso como o Brasil, a utilização de soluções próprias nos municípios pode acarretar diversos problemas para a gestão desses dados pelos estados e União. Devido a possíveis dificuldades de integração dos controles próprios com o SINAN, existe o risco da demora no envio desses dados, ocasionando problemas na consolidação dos dados e no monitoramento ativo da situação epidemiológica em cada localidade. Como a questão temporal é fundamental em doenças como a dengue, esse atraso pode gerar impactos nos processos de tomada de decisão.

Outro possível problema reside no fato de que o desenvolvimento de soluções próprias normalmente não segue um mesmo padrão e elas podem não estar adequadas às diretrizes nacionais, aumentando ainda mais as dificuldades da integração com outros sistemas, seja por barreiras técnicas ou inadequação em relação aos protocolos definidos pelo Ministério da Saúde. Além disso, é importante observar que os sistemas locais podem não ser aderentes a necessidades básicas de segurança para proteção dos dados dos pacientes. Em se tratando de dados tão sensíveis como os relacionados a doenças, a falta de segurança pode comprometer a própria confiança dos usuários no SUS.

Existem outros sistemas relacionados à dengue que não estão necessariamente integrados, como os sistemas para gestão dos dados laboratoriais e os utilizados para controle das atividades realizadas pelas equipes de saúde. Dessa forma, torna-se mais difícil a avaliação contínua da eficácia das intervenções. Com a possível falta de

integração automática dos diversos sistemas existentes, esse processo muitas vezes acaba sendo realizado através da digitação dos dados. Além da questão da velocidade de preenchimento, que é sujeita a disponibilidade de recursos humanos, há ainda riscos de erros nessa transposição manual de dados entre sistemas, o que pode comprometer ainda mais as análises dos dados para a tomada de decisão.

2.2 Avaliação em Saúde

A avaliação de resultados é etapa fundamental no ciclo de desenvolvimento de qualquer projeto, visto que é a forma de medir se as atividades estão cumprindo os resultados esperados. Nessa etapa, os resultados podem ser avaliados através de um ou mais indicadores, que permitem aos gestores a identificação de pontos de melhoria. Com isso, adaptações ou até mudanças mais profundas são realizadas e o ciclo recomeça. Existem diversos modelos de melhoria contínua e um dos mais conhecidos é o ciclo PDCA (*plan-do-check-act*), em que a avaliação é feita na terceira etapa (*check*) e serve como base para a ação (ex: ajustes nas atividades) para a próxima interação do ciclo (Figura 7).

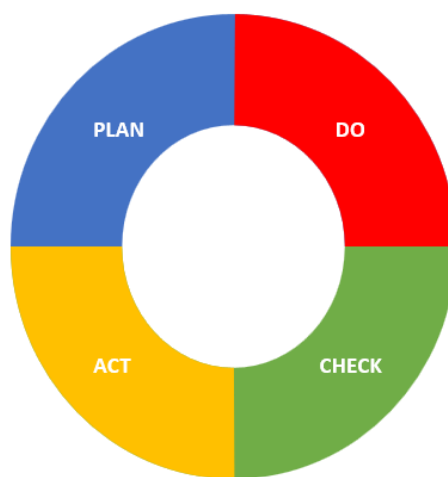


Figura 7 - Ciclo PDCA.

Na área da saúde, a avaliação é essencial para verificar se as intervenções estão gerando os efeitos esperados, como a diminuição de casos de uma determinada doença ou a melhora das condições clínicas dos pacientes. O desenvolvimento de medicamentos é um exemplo cujo processo pode levar anos até que o medicamento

seja disponibilizado no mercado. Para as políticas de saúde, a avaliação dos resultados pode indicar diversos aspectos, como:

- As intervenções são adequadas para gerar os efeitos esperados?
- As ações estão sendo executadas conforme preconizado?
- Quais as condições necessárias para execução das ações?

Logo, o processo de avaliação serve como base para a tomada de decisão pelos órgãos responsáveis, como adaptar, continuar ou até interromper a execução da intervenção. Além disso, como os cenários mudam, a avaliação pode orientar a gestão pública se determinada intervenção ainda é adequada no novo contexto.

2.2.1 Tipos de Avaliação

Como a avaliação pode servir a diferentes objetivos, a avaliação em saúde pode ser dividida em dois tipos: pesquisa avaliativa e avaliação normativa (CONTANDRIOPOULOS et al., 2000). As avaliações do primeiro tipo têm como objetivo a geração de conhecimento sobre um assunto, como o desenvolvimento de uma forma de tratamento para uma doença, por exemplo. Já as avaliações do tipo normativa buscam identificar se as atividades estão sendo desenvolvidas conforme critérios e normas estabelecidas. Enquanto a pesquisa investigativa tem um caráter mais relacionado à pesquisa científica, as avaliações normativas são mais relacionadas a processos de gestão. Nada impede, no entanto, que uma mesma intervenção seja objeto dos dois tipos de avaliação.

A avaliação pode considerar os diversos componentes da intervenção, visto que a adoção de um tratamento, por exemplo, deve considerar outros fatores além do possível medicamento em si. Na Figura 8, são ilustrados os cinco componentes de uma intervenção e observa-se a complexidade envolvida na avaliação de uma intervenção. Dada uma determinada situação (ex: surto de uma doença), são definidos objetivos (ex.: criação de uma política) que utilizam recursos e serviços (ex: tratamentos, profissionais, unidades de saúde etc.) para a geração dos efeitos esperados (ex.: diminuição do número de casos). Esses quatro componentes ainda estão inseridos em um contexto específico que pode ter influência em todos os demais componentes.

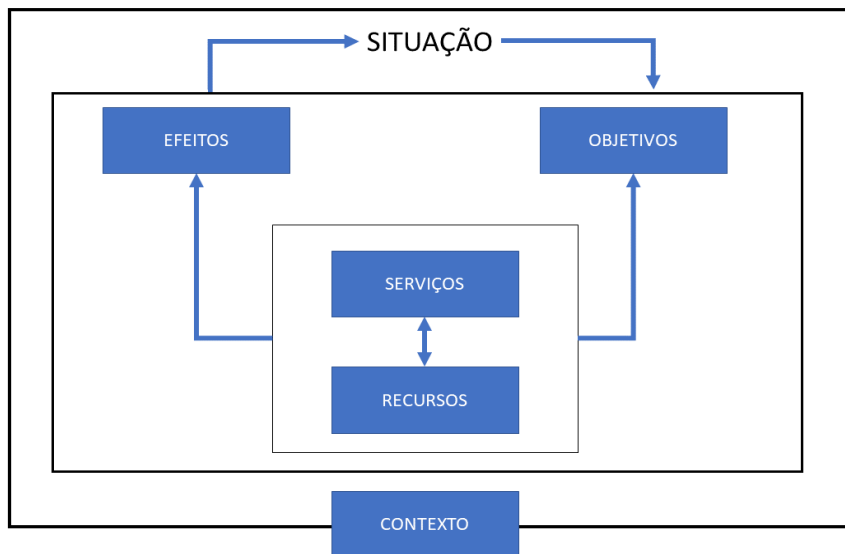


Figura 8 - Componentes da intervenção.

Fonte: Adaptado de (CONTANDRIOPOULOS et al., 2000).

Logo, um processo avaliativo de uma determinada intervenção envolve mais do que avaliação da eficácia de uma ação em si, pois está relacionada a serviços e recursos e inserida em um contexto, com o qual pode ter graus diferentes de interação dependendo do aspecto analisado. Por exemplo, no combate à dengue, as ações de controle vetorial poderiam gerar mudanças no contexto, como a diminuição dos índices de infestação. Em relação ao clima, essas ações podem ser intensificadas no caso de um cenário que favoreça à proliferação dos vetores, como épocas de maior temperatura.

Nos dois tipos de avaliação, os componentes da intervenção podem ter avaliações específicas. No caso da avaliação normativa, a avaliação pode ser feita para os componentes recursos, serviços e efeitos de acordo com normas e critérios estabelecidos. Quanto aos recursos, a infraestrutura utilizada pode ser objeto de avaliação, por exemplo. A quantidade de unidades de saúde, leitos ou pessoal pode ser verificada quanto às normas existentes. Por exemplo, de acordo com as diretrizes de combate à dengue, deve haver um agente comunitário de saúde para no máximo 750 habitantes. Logo, os resultados da intervenção podem ser impactados se essa norma não for atendida por algum município.

Em relação aos serviços, podem ser avaliados conforme indicadores definidos nas normas. As diretrizes de combate à dengue indicam que devem ser feitas visitas aos imóveis especiais com certa periodicidade para verificação e controle de possíveis criadouros. Assim como em relação aos recursos, a ausência desse serviço pode ter impacto nos índices de infestação e, conseqüentemente, no número de casos de dengue.

A avaliação do componente “Efeitos” tem como objetivo verificar se os resultados alcançados correspondem ao esperado. Pode-se identificar, no caso do controle vetorial, que as ações não surtiram o efeito esperado na diminuição dos índices de infestação. No entanto, as razões pelas quais esses resultados foram insatisfatórios podem não ser respondidas apenas por esse tipo de avaliação. Nesse sentido, pode ser conduzida a pesquisa avaliativa de forma complementar a avaliação normativa. Esse tipo de avaliação consiste na aplicação do método científico para realizar o julgamento da uma intervenção. Conforme Contandriopoulos et al. (2000), esse tipo pode ser decomposto em seis tipos de análise, conforme Figura 9.

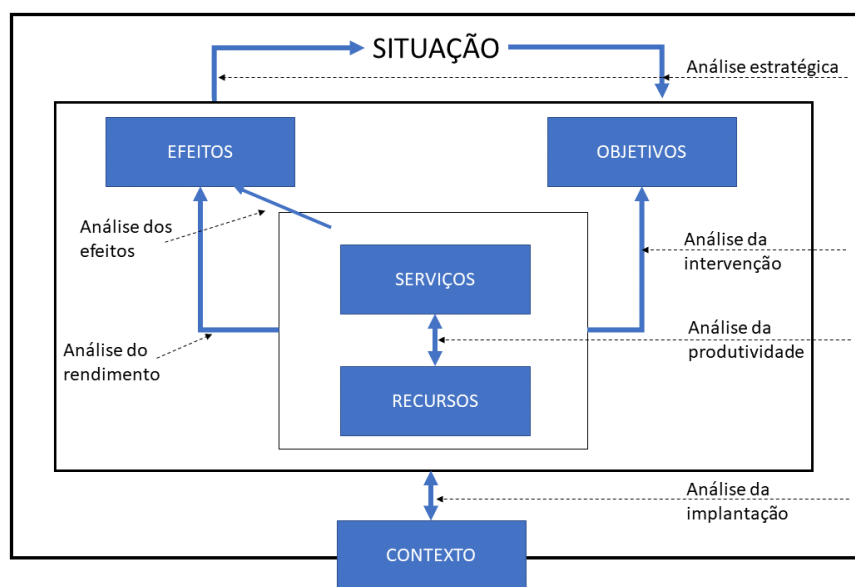


Figura 9 - Tipos de análise na pesquisa avaliativa.

Fonte: CONTANDRIOPOULOS *et al.*(2000, adaptação nossa).

A **Análise Estratégica** consiste em avaliar a adequação estratégica entre a intervenção adotada e a situação (ex.: infestação de mosquitos), ou seja, se é pertinente

intervir nesse problema considerando outros problemas existentes. Também pode ser avaliada a forma da intervenção adotada, assim como analisar se o fator de risco no qual se quer intervir é o mais importante. Na **Análise da Intervenção** busca-se avaliar a adequação dos recursos e serviços frente aos objetivos definidos, tanto na forma quanto na quantidade. Pode-se avaliar, por exemplo, se os produtos adotados para controle de criadouros estão capazes de impedir a proliferação dos mosquitos.

A relação entre os recursos e serviços é objeto da **Análise da Produtividade**, na qual pode-se avaliar a produção dos serviços quanto aos recursos disponíveis. Questões podem surgir como, *poderiam ser produzidos mais serviços com os mesmos recursos?* Esse tipo de avaliação pode gerar respostas sobre a capacidade dos recursos para geração dos serviços necessários para a intervenção. Já a **Análise dos Efeitos** consiste em avaliar a influência dos serviços adotados nos estados de saúde, ou seja, na determinação do impacto dos serviços para modificação dos estados de saúde. A questão da eficiência é objeto da **Análise do Rendimento**, que consiste na avaliação da relação entre os recursos e os efeitos obtidos. Nesse tipo, podem ser realizadas análises custo/benefício ou custo/eficácia, considerando indicadores como redução de números de casos, redução da letalidade etc.

Por fim, a **Análise da Implantação** consiste em medir a influência que o grau de implantação pode ter nos efeitos, assim como possíveis impactos do ambiente nos resultados. Esse tipo de análise pode ser adotado quando se observa grande variabilidade nos resultados das mesmas intervenções em contextos diferentes. Considerando o Programa de Combate à Dengue, poderiam ser avaliadas possíveis diferenças dos resultados de controle vetorial em dois municípios diferentes, por exemplo. Além disso, pode-se avaliar a contribuição do ambiente para os efeitos de uma intervenção, tanto positiva quanto negativamente, ou seja, se o ambiente teria um efeito inibidor na intervenção.

Esses tipos de análises podem ser adotados na sua totalidade quanto por partes. Pelo fato de atenderem a *stakeholders* diferentes, a adoção de cada tipo de análise pode estar relacionada à origem da demanda. Por exemplo, a área responsável pelo planejamento do orçamento de um município poderia utilizar a análise da quantidade

de recursos necessários para combate à dengue considerando a produtividade em um cenário de epidemia previsto.

Uma outra abordagem baseou-se em oito critérios para identificar os tipos de avaliação em saúde: pesquisa de avaliação, avaliação para decisão e avaliação para decisão (NOVAES, 2000). Conforme descrito na Tabela 8, os tipos podem variar quanto a diversos critérios, como objetivo da análise, tipo de juízo formulado, entre outros. As avaliações de cada tipo podem ser conduzidas individualmente ou em conjunto, e podem ser aplicadas em avaliações de programas de saúde ou em avaliações de tecnologias em saúde.

Tabela 8: Descrição dos tipos de avaliação.

Critério	Pesquisa de avaliação	Avaliação para decisão	Avaliação para gestão
Objetivo	Conhecimento	Tomada de decisão	Aprimoramentos
Posição do avaliador	Externo ou interno	Interno ou externo	Interno ou externo
Enfoque priorizado	Impactos	Caracterização ou compreensão	Caracterização ou quantificação
Metodologia dominante	Quantitativo (qualitativo) experimental ou quasi-experimental	Qualitativo e quantitativo situacional	Quantitativo e qualitativo situacional
Contexto da avaliação	Controlado	Natural	Natural
Forma de Utilização da informação	Demonstração	Informação	Instrumentos para gestão
Tipo de juízo formulado	Hipóteses	Recomendações	Normas
Temporalidade da avaliação	Pontual/replicado	Corrente/pontual	Integrado/contínuo

Fonte: (NOVAES, 2000)

O principal critério de diferenciação é o objetivo, ou seja, o que deve ser produzido pelo processo avaliativo. Na **Pesquisa de Avaliação**, define-se como principal objetivo a geração de conhecimento em relação ao objeto, por exemplo, o desenvolvimento de uma intervenção. Na **Avaliação para Decisão** busca-se gerar meios de suporte aos processos decisórios, como a geração de recomendações sobre quais intervenções devem ser adotadas. Os processos de melhoria contínua descritos anteriormente aproximam-se do tipo **Avaliação para Gestão**, que tem como objetivo

a produção de informação (por exemplo, indicadores) para melhoria do objetivo avaliado.

Na avaliação de programas de saúde, como o programa de combate à dengue, podem ser aplicadas avaliações dos três tipos. Os tipos de avaliações podem ter diferentes graus de aderência aos seis tipos de análises dos componentes de uma intervenção descritos anteriormente. As análises estratégicas e de implantação podem estar mais relacionadas às avaliações para decisão, enquanto a análise de rendimento está mais conectada a à avaliação para decisão. A pesquisa avaliativa, por sua vez, pode estar mais associada a análise de efeitos. No entanto, nada impede que esse tipo de avaliação seja aplicado a outro tipo de análise, como a análise de implantação, caso seja de interesse conhecer os impactos que o contexto pode gerar nos resultados das intervenções.

Além disso, os resultados de cada tipo de pesquisa podem ser usados nos outros tipos, com os resultados de um tipo serem base para os demais. Por exemplo, uma pesquisa de avaliação pode identificar diversos produtos para controle de criadouros de mosquitos, informação essa que pode ser usada na avaliação para decisão para recomendar qual produto adotar em cada contexto. A avaliação para gestão poderia utilizar-se de indicadores para medir a utilização de cada produto e, dessa forma, gerar informações para uma análise retrospectiva de uma pesquisa de avaliação. Embora os três tipos de avaliação possam ser adotados isoladamente, essa possível integração (Figura 10), entre as avaliações dos três tipos poderia agregar valor aos programas de saúde.

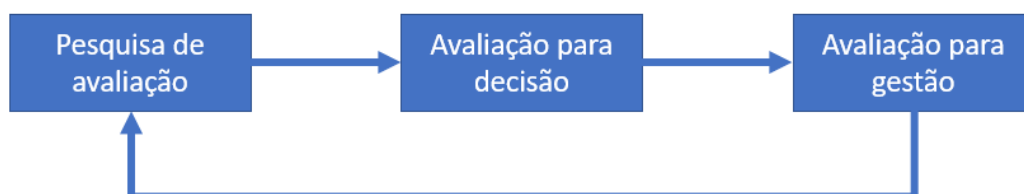


Figura 10 - Uso integrado dos tipos de avaliação.

2.2.2 Modelo de Avaliação de Donabedian

Como a questão da saúde está além da prática médica em si, os processos de avaliação devem considerar outros componentes e suas interconexões. Uma das abordagens mais adotadas para avaliação em saúde é o modelo de Avedis Donabedian (DONABEDIAN, 1988), no qual são considerados indicadores de três categorias: estrutura, processo e resultado, conforme Figura 11.

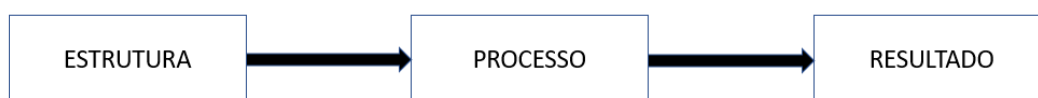


Figura 11 – Componentes do modelo de avaliação de Donabedian.

Fonte: DONABEDIAN(1988, adaptação nossa).

Para Donabedian (1988), a categoria **Estrutura** compreende as condições em que o cuidado médico é aplicado, como instalações, equipamentos e recursos humanos. O **Processo** refere-se às atividades como busca pelos serviços até as atividades de diagnóstico e tratamento. Os efeitos resultantes no estado de saúde do paciente ou população referem-se à categoria **Resultado**. A avaliação baseada nessa tríade considera que uma estrutura adequada aumenta a probabilidade de processos de qualidade, ao passo que esses aumentam a probabilidade de bons resultados.

No caso do programa de combate à dengue, o modelo de Donabedian poderia ser utilizado de várias formas. Poderiam ser considerados, por exemplo, o índice de infestação e a incidência de casos como variáveis a serem avaliadas no componente **Resultado**. As ações preconizadas no PNCD, como o controle de criadouros, poderiam ser objetos para o componente **Processo**, enquanto que os recursos humanos (ex.: número de agentes de saúde) seriam variáveis no componente **Estrutura**.

2.3 Big Data e Ciência de Dados

Com a maior disponibilidade de fontes de dados e aumento da capacidade de processamento computacional, termos como *Big Data*, *Ciência de Dados* e *Machine Learning* tem recebido cada vez mais destaque tanto na área da pesquisa em Computação quanto no mercado de trabalho para novas profissões (AGARWAL;

DHAR, 2014; VAN DER AALST, 2014). O relatório do Fórum Econômico Mundial de 2020 apontou que Cientistas de Dados e Especialistas em Inteligência Artificial estão entre as profissões com maiores taxas de crescimento em empregos (WORLD ECONOMIC FORUM, 2020).

Ao mesmo tempo que a era do *Big Data* apresenta diversas oportunidades para descoberta de conhecimento, também há desafios que envolvem questões como eficiência para extração de informação, privacidade, entre outras. Entre as possíveis abordagens para lidar com esse cenário está a Ciência de Dados, área que contém características e técnicas que podem ser aderentes à pesquisa em Saúde Planetária. O uso de técnicas de aprendizado de máquina, por exemplo, já tem grande aplicação em pesquisas na área da saúde.

2.3.1 Oportunidades e Desafios

Com a profusão de dispositivos com capacidade computacional e o aumento no acesso à Internet, cada vez mais dados são gerados. Essa era, conhecida como *Big Data*, também é caracterizada pela diversidade dos formatos de dados, assim como pela velocidade com que são gerados (SAGIROGLU; SINANC, 2013). Além de texto, dados em outros formatos, como multimídia, são gerados a cada instante por vários dispositivos, especialmente os *smartphones*. A maior interoperabilidade dos sistemas computacionais estimulada, entre outras coisas, por um mundo cada vez mais conectado, também contribui para que esse volume de dados seja ainda maior. Sistemas informatizados geram dados que são consumidos por outros sistemas que, por sua vez, geram respostas num processo de comunicação intensificado pela Internet, contribuindo para esse dilúvio de dados.

Ao contrário de décadas atrás, em que o volume de dados era concentrado em dados estruturados, atualmente há predominância dos dados não-estruturados (ADNAN; AKBAR, 2019), cujo domínio é estimulado principalmente pelas redes sociais, outro reflexo da maior disponibilidade de dispositivos conectados à Internet e de uma sociedade cada vez mais conectada. Todo cidadão conectado por meio de *smartphones* produz conteúdo e isso não está mais restrito aos grandes portais de notícias. Através de plataformas como redes sociais, *blogs* e ambientes de

compartilhamento multimídia, dados são gerados de maneira muito rápida e esses ambientes são, de certa forma, boa parte da fonte de informação dos portais de notícias, invertendo a ordem de duas décadas atrás.

Além dos dados compartilhados espontaneamente pelos usuários, ainda há aqueles que são gerados de forma passiva. Qualquer *smartphone* tem sensores de localização embutidos, dispositivo que pode gerar dados de mobilidade do usuário sem que o mesmo muitas vezes se dê conta disso. Plataformas podem fazer uso desses dados para oferecer serviços e produtos ao usuário e muitas delas armazenam dados de localização do usuário estando o mesmo conectado ou não a essas plataformas. Existem diversas questões relacionadas à segurança e privacidade dos usuários, temas que a Lei Geral de Proteção de Dados (LGPD) do Brasil busca tratar (BRASIL, 2018b).

Com um custo cada vez menor de armazenamento, e com os métodos de compartilhamento facilitados pelos recursos computacionais em nuvem, armazenar dados não é o maior dos desafios na era do *Big Data*. Custos de *hardware* para processamento de dados também não chega a ser um desafio, dado o seu barateamento e desenvolvimento cada vez maior da indústria de microeletrônicos. Embora a quantidade imensa desses dispositivos de processamento e armazenamento de dados possa ser questionada do ponto de vista ambiental, já que esses recursos consomem energia e geram problemas do ponto de vista de gestão dos resíduos pela obsolescência rápida das máquinas e necessidade de trocas por outras, não se pode afirmar que esse seja o maior desafio da era do *Big Data*, mesmo sendo muito importante na busca de uma sociedade mais sustentável.

O que move boa parte das pesquisas em Computação relacionadas à *Big Data* reside em (a) o que fazer com esses dados? e (b) como extrair informação útil de maneira eficiente?

Não basta ter um volume imenso de dados se não há uso deles. Dados brutos só são úteis quando transformados em informações para geração de *insights* e conhecimento. Considerando, além do volume, a diversidade de formatos e fontes de dados, métodos outrora úteis para análise de dados não são mais adequados isoladamente no contexto do *Big Data*. Outras abordagens são necessárias para

extração de informação na nova realidade mundial em que “*data is the new oil*”, numa analogia que não se refere apenas ao valor dos dados, mas também relacionando o processo de extração de dados com a extração de petróleo (VAN DER AALST, 2014).

Mas não basta apenas extrair a informação a partir desse volume de dados. A forma com que se faz isso também é muito importante e objeto de muitas pesquisas pois, uma vez que os dados são gerados em velocidade cada vez maior, uma informação não extraída eficientemente pode tornar-se obsoleta rapidamente. Logo, a velocidade com que se extrai informação é fator primordial no contexto de *Big Data*.

Dessa forma, diversos métodos computacionais têm ganhado ainda mais importância no mundo do *Big Data*, com destaque para aqueles da área de Inteligência Artificial (IA). Mas por que IA tem recebido tanta atenção, tanto no mercado quanto no ambiente de pesquisa? Alguns fatores podem explicar esse fato, mas destacam-se dois: 1) maior capacidade atual de processamento computacional e, 2) inadequabilidade das abordagens tradicionais de análise de dados no mundo de *Big Data*. A necessidade de extrair informação de forma eficiente associada ao barateamento dos recursos de processamento computacional constitui em grande estímulo para o desenvolvimento da pesquisa em IA, com o investimento de grandes corporações tanto quanto a criação de linhas e centros de pesquisas relacionadas no ambiente acadêmico.

No entanto, apenas os métodos de IA não são suficientes no contexto de *Big Data*. Considerando a diversidade de fontes de dados, cada qual com seus formatos e meios de integração próprios, o processo de coleta e processamento de dados também exige esforço adicional na disponibilização dos dados para o uso dos métodos de IA. Nesse sentido, todo o conhecimento gerado em disciplinas relacionadas a padrões de dados, metadados e na integração desses tem ainda mais importância na gestão de grandes repositórios de dados oriundos de diversas (e de formatos variados) fontes. Métodos provenientes de *Business Intelligence* (BI) são igualmente úteis nesse cenário. Sistemas de BI, que buscam agregar ferramentas analíticas a dados operacionais, existem há muitos anos e podem prover informação útil para gestores no processo de tomada de decisão (NEGASH, 2004).

Outras áreas da Computação também são importantes considerando o cenário de *Big Data* pois, mesmo com o aumento do poder computacional, existem limites que são facilmente superados pelo volume de dados dos grandes repositórios. Na busca de extração de informação útil de maneira eficiente, a integração e os melhores métodos de IA podem ainda não ser suficientes. Por isso, métodos de processamento paralelo e computação distribuída muitas vezes são usados para análise de dados, e a era do *Big Data* tem mostrado ainda mais a importância desses métodos.

Além de BI, destaca-se o *Data Warehouse* (DW) que é, segundo Inmon, “uma coleção de dados orientada por assunto, integrada, variante no tempo e não volátil para apoiar o processo de tomada de decisão da administração” (INMON, 2005). Os DWs foram fundamentais para o avanço da visão analítica dos dados transacionais e não deixaram de ser usados no cenário de *Big Data*. No entanto, os ambientes de DW tradicionais mostraram-se insuficientes para atender às demandas inerentes ao cenário com massivos e variados volumes de dados. Nesse sentido, abordagens alternativas vêm sendo propostas para gestão dos dados para atividades analíticas. Uma delas é o *Data Lake* (DL), que pode ser definida como um vasto repositório que armazena dados em seus formatos originais (NAMBIAR; MUNDRA, 2022).

Como um mesmo conjunto de dados pode servir a diferentes propósitos, a manutenção dos dados em seus formatos originais transfere a decisão de quais transformações aplicar para o momento em que os dados são utilizados. Diferente dos processos tradicionais de análise de dados nos *Data Warehouses* (DW), em que o *pipeline* de dados consiste normalmente no modelo ETL (*extract-transform-load*), nas abordagens baseadas em *Data Lake* (DL), a etapa de transformação é executada após o carregamento dos dados, ou seja, em um modelo *extract-load-transform* (ELT). Outra diferença importante é que em um DW os dados estão vinculados a um esquema pré-definido, no modelo chamado *schema-on-write*. Já nos DLs o esquema pode ser definido no uso dos dados (*schema-on-read*), permitindo maior flexibilidade e velocidade para ingestão de conjuntos de dados no repositório.

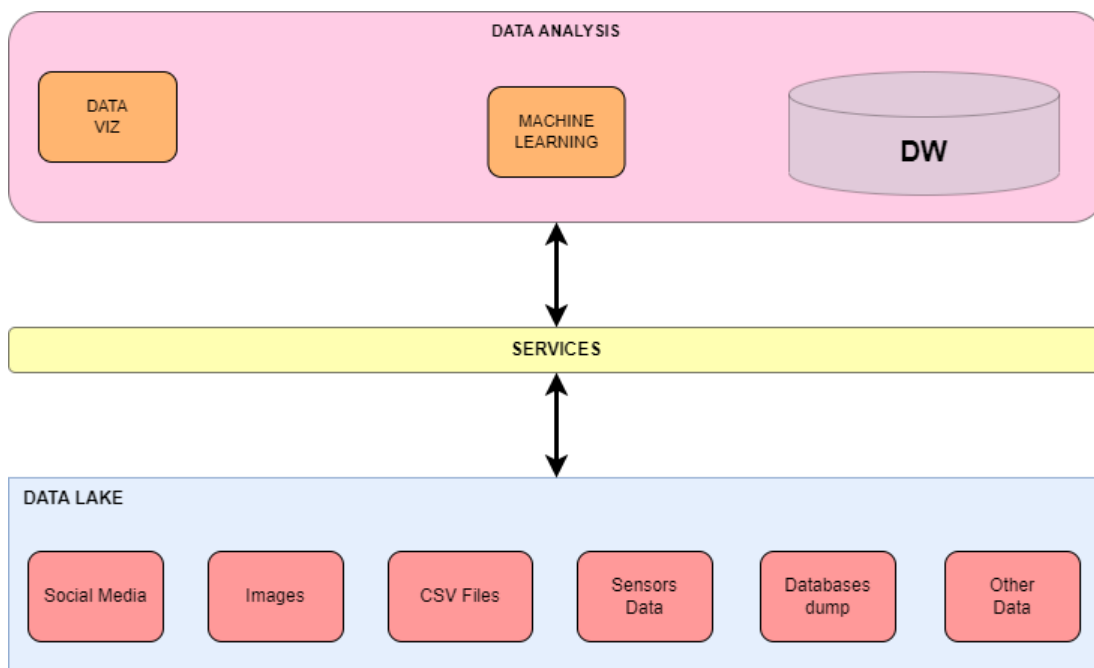


Figura 12 - Exemplo de estruturação de um *Data Lake*.

Com um *Data Lake*, as atividades de análise de dados podem buscar e selecionar os conjuntos de dados necessários, que podem ser de diversos formatos e fontes, conforme Figura 12. Um DL não elimina necessariamente os *Data Warehouses* que, na verdade, podem coexistir com os DL e sendo uma forma de integração de dados do DL para realização de análises para um determinado objetivo. Existem muitas arquiteturas possíveis para os DLs, que podem variar de acordo com o contexto em que são aplicados. No exemplo da Figura 12, pode existir uma camada intermediária de serviços (*services*) para descoberta e seleção de conjuntos de dados. Outra abordagem é o *Lakehouse*, que busca combinar os benefícios dos *Data Warehouses* com os dos *Data Lakes* (ARMBRUST1 et al., 2021) como ilustrado na Figura 13.



Figura 13 – Figura ilustrativa de um Lakehouse e suas diferenças para outros padrões arquiteturais de dados.

Fonte: (LORICA et al., 2020)

Embora se diga que abordagens computacionais como aprendizado de máquina e Data Lakes sejam fundamentais no mundo do *Big Data*, não se pode afirmar que essas abordagens substituem as técnicas tradicionais de análise de dados, muito pelo contrário. Esse novo contexto traz a necessidade de integração das áreas de conhecimento e, nesse sentido, abordagens que integram áreas como Matemática, Estatística e Computação têm sido adotadas para análise de dados de grandes e diversos volumes de dados. Isso se reflete na própria formação acadêmica, visto que há cada vez mais necessidade de se formar engenheiros e cientistas de dados, e não apenas cientistas da computação, matemáticos e estatísticos.

No entanto, apenas as áreas relacionadas a análise de dados não são suficientes nesse cenário. O conhecimento a respeito dos dados, tanto na geração e coleta deles quanto no domínio que estão inseridos, é também muito importante. Conhecer os dados e determinar que perguntas devem ser feitas sobre esses dados é uma competência vinda principalmente dos especialistas de domínio. São esses que devem apresentar os problemas a serem resolvidos pelos especialistas em análise de dados, além de participar ativamente do processo de avaliação dos resultados preliminares bem como dos produtos finais. O processo de extração de informação é baseado muitas vezes em ciclos de alta interação entre os especialistas de domínio e os analistas de

dados. Nesse sentido, abordagens como Ciência de Dados têm sido cada vez mais aplicadas em projetos relacionados à *Big Data*.

2.3.2 Ciência de Dados e Saúde Planetária

As respostas para as duas questões colocadas anteriormente em relação ao que move as pesquisas em *Big Data*, passa por a) desenvolvimento das áreas do conhecimento apontadas nos parágrafos anteriores e b) a integração de diversas áreas do conhecimento para extração de informação e geração do conhecimento. Para essa segunda, destaca-se a Ciência de Dados (em inglês, *Data Science*) que é uma das áreas com maior destaque nos últimos anos.

Existem diversas discussões se *Data Science* é ou não uma área nova, mas não se pretende entrar no mérito dessa questão aqui. Enquanto alguns pesquisadores podem se referir a *Data Science* como “*an old wine in a new bottle*”, outros tratam-na como uma nova área da Estatística ou, ainda, uma nova ciência (CLEVELAND, 2001). No entanto, o que importa aqui é ressaltar o seu caráter multidisciplinar e a sua importância entre as abordagens de destaque na extração eficiente de informação útil no mundo do *Big Data*. Não se pode esperar também, dada a complexidade de áreas envolvidas, que um profissional tenha todos os conhecimentos desse “guarda-chuva” de áreas. Na verdade, um projeto de *Data Science* envolve muito mais do que um profissional com habilidades em uma ou duas áreas, mas sim a atuação de um time multidisciplinar, compostos por profissionais com conhecimentos em Computação, Estatística, além dos especialistas de domínio, que podem ter formação nessas ou em outras áreas do conhecimento.

Quando se analisa um problema na área de Saúde Planetária, por exemplo, há características que demandam uma abordagem multidisciplinar. Algumas das características relativas aos dados no contexto de Saúde Planetária são bem aderentes ao que convencionou-se chamar de *Big Data*: multiplicidade de fontes de dados, alto volume e grande velocidade de geração. Considerando-se ainda que os determinantes da saúde humana são relacionados a diversas áreas, como meio ambiente, alimentação e economia, vê-se que a abordagem para pesquisas em Saúde Planetária é, naturalmente, multidisciplinar e transdisciplinar. Portanto, a análise de dados no

contexto de um problema de Saúde Planetária, como a dengue, também pode ser abordada pela proposta de integração de áreas como em projetos de *Data Science*.

2.3.3 Big Data e Aprendizado de Máquina

Uma das técnicas utilizadas em projetos de Ciência de Dados é o aprendizado de máquina (em inglês, *machine learning* - *ML*), um processo computacional em que algoritmos são utilizados para aprender informações a partir de conjuntos de dados, como reconhecimento de padrões e geração de modelos preditivos (EL NAQA; MURPHY, 2015).

Embora ML não seja necessariamente uma técnica nova, o advento do Big Data apresenta diversos desafios para o desenvolvimento de pesquisas em ML (L'HEUREUX et al., 2017). A expansão no uso de ML acontece em vários níveis, como o desenvolvimento de novos algoritmos, além de formas de armazenamento e processamento dos dados. Além disso, dada a necessidade de ter respostas rápidas frente às mudanças de cenário, os processos de desenvolvimento de algoritmos têm incorporado atividades para gestão do ciclo concepção-desenvolvimento-implantação-melhoria dos modelos de aprendizado de máquina (TREVEIL et al., 2020).

Somado a isso, têm sido desenvolvidas abordagens para gestão e integração dos dados, considerando as características de cada área e as demandas para produção e melhoria dos algoritmos. O acesso a grandes e diversos volumes de dados de maneira eficiente pode aumentar o potencial de predição dos modelos, assim como suas adaptações face a novos cenários. Para isso, abordagens relacionadas aos *data lakes* têm ganhado atenção, possibilitando a incorporação de diversas fontes de dados para serem utilizados pelos modelos de ML (MILOSLAVSKAYA; TOLSTOY, 2016).

Outro fator refere-se à privacidade das pessoas, tema que é ainda mais essencial em áreas com dados sensíveis, como a saúde. Ao mesmo tempo que a descoberta de conjuntos de dados úteis demanda meios eficientes, a proteção desses dados desde a sua ingestão nas estruturas de armazenamento deve ser garantida de modo a garantir os direitos dos indivíduos. Além disso, dada a maior incorporação dos algoritmos nas atividades, há aumento da necessidade do desenvolvimento de mecanismos para trazer transparência aos modelos (BÜCKER et al., 2022).

2.3.4 Aplicações de ML na Saúde

Embora esses fatores e desafios sejam inerentes à adoção de ML em qualquer área, pode haver especificidades que podem determinar as prioridades para o desenvolvimento de pesquisas. Na saúde, uma taxonomia foi criada para organizar as aplicações de aprendizado de máquina de acordo com as prioridades de pesquisa nesse campo (MHASAWADE; ZHAO; CHUNARA, 2021):

- **Identificação dos fatores e suas relações com a saúde.** ML pode ser usada para aprender que elementos na estrutura socioecológica contribuem para os resultados da saúde. Embora a pesquisa já tenha tido grande avanço na identificação de diversos determinantes da saúde, a incorporação de novas fontes de dados aliada à capacidade de processamento de aprendizado de máquina pode trazer novos conhecimentos, como identificação de mais fatores além de melhor compreensão da relação multinível dos fatores com as doenças;
- **Desenvolvimento de intervenções.** Considerando a relação de diversos fatores com as doenças, há a necessidade de intervenções em vários níveis. Assim, ML pode ser usada para avaliar a eficácia dessas intervenções de acordo com suas aplicações nos diversos níveis. Além disso, considerando que os locais não são iguais, a análise das intervenções poderia considerar as especificidades de cada local onde as intervenções são aplicadas, o que ainda é mais necessário em um país tão diverso como o Brasil;
- **Previsão de desfechos.** Existem grandes avanços nessa área, especialmente nos eventos dentro das unidades de saúde. A incorporação de novos fatores e o desenvolvimento de novos modelos pode expandir essas análises para outras áreas da saúde, como aquelas que envolvem a saúde pública. Além disso, a integração de outras variáveis podem ser úteis para melhorar a prática médica de acordo com os fatores de risco de seus pacientes. A aplicação de ML nesse tema poderia considerar questões relativas à desigualdade e como isso pode impactar nos desfechos de saúde;
- **Alocação de recursos.** O uso de ML pode auxiliar nos processos de tomada de decisão, servindo como base para estabelecer prioridades de alocação dos

recursos. Pode considerar questões como custo-efetividade das intervenções, onde a variável custo pode ir além de questões econômicas. No combate à dengue, por exemplo, o uso de controle químico pode ter um custo ambiental importante que deve ser avaliado frente a alternativas de controle vetorial.

O cenário de *Big Data*, ao mesmo tempo que apresenta diversos desafios principalmente para as áreas da Computação, traz oportunidades para desenvolvimento de conhecimento em diversos setores. Na saúde, a aplicação dos métodos computacionais a grandes conjuntos de dados pode agregar grande conhecimento na área, como desenvolvimento de intervenções e prover respostas rápidas em cenários em que a velocidade é fundamental, como epidemias.

Para doenças como a dengue, em que são múltiplos fatores envolvidos, os métodos relacionados a *Big Data* podem auxiliar a gestão e pesquisa na área na descoberta e avaliação das intervenções. Isso pode ser tornar ainda mais necessário considerando os cenários em que as condições para disseminação dos vetores serão ainda mais favoráveis, como as mudanças climáticas e o aumento da urbanização.

2.3.5 Questões emergentes sobre dados

Com o advento da era de *Big Data*, as discussões sobre a gestão dos dados vão além de questões técnicas computacionais. A maior disponibilidade dos dados, ao mesmo tempo que representa oportunidade para avanço no conhecimento em diversas áreas, também faz com que outras questões sejam aprofundadas.

Uma dessas questões refere-se ao reuso dos dados, que pode ser fundamental para o desenvolvimento de pesquisas colaborativas. Os princípios FAIR (*Findable, Accessible, Interoperable, Reusable*) descrevem pontos essenciais que poderiam ser considerados nos projetos de dados visando a colaboração (WILKINSON et al., 2016). Esses princípios têm como foco o aprimoramento dos sistemas para busca e uso automático dos dados e apoio ao reuso por todos os interessados. Para serem utilizados, os dados precisam ser encontráveis (*Findable*), o que pode ser apoiado com uma gestão eficiente de metadados que permita que conjuntos de dados úteis possam ser encontrados. Além disso, os dados devem ser acessíveis (*Accessible*), recursos que podem ser providos pelos sistemas através de mecanismos de autenticação e

autorização. A questão da interoperabilidade (*Interoperable*) também é fundamental para o reuso, visto que os dados normalmente precisam ser integrados a outros dados. Por fim, informações podem ser adicionadas aos metadados para otimizar o reuso (*Reusable*) dos dados, como licenças de uso, proveniência etc.

Os princípios FAIR estão relacionados a outra questão fundamental: dados abertos. Embora esses princípios não impliquem necessariamente que os dados sejam abertos para reutilização, a publicação de dados de forma pública poderia considerar os princípios FAIR para otimizar a sua reutilização. Dados abertos podem ser muito úteis em diversas áreas, como a saúde, em que dados médicos compartilhados podem auxiliar as pesquisas para melhoria dos sistemas de saúde e tratamento de pacientes (KOSTKOVA et al., 2016).

No entanto, o compartilhamento dos dados usando os princípios FAIR e, quando possível, de forma aberta, leva a outro desafio: a proteção dos direitos individuais. Nesse sentido, para proteção desses direitos, foi criada no Brasil a Lei Geral de Proteção de Dados para garantir a gestão segura de informações pessoais (RANK; BERBERI, 2022). Assim, as instituições devem prover mecanismos para garantir a proteção de direitos como privacidade e segurança. Técnicas como anonimização são normalmente utilizadas para proteger informações sensíveis, como dados provenientes de hospitais e instituições financeiras (MAJEED; LEE, 2021).

A maior disponibilidade de dados, ocasionada pelo advento de Big Data, associada ao compartilhamento desses dados entre instituições e indivíduos, pode potencializar os ganhos das atividades de descoberta de conhecimento. Nesse sentido, diversas iniciativas relacionadas ao compartilhamento de dados surgiram nos últimos anos.

Uma delas é a *Open Government Partnership* (OGP), da qual o Brasil é um dos oito países fundadores, e tem como objetivo a difusão de práticas para transparência, acesso à informação e participação social (BRELÀZ; CRANTSCHANINOV; BELLIX, 2021). Os países devem assumir compromissos relacionados a quatro princípios: transparência, participação cidadã, prestação de contas e tecnologia e inovação (CONTROLADORIA-GERAL DA UNIÃO, 2023).

Para os anos de 2021 a 2023, o Brasil desenvolveu um plano com 12 compromissos, contendo objetivos e ações para o cumprimento das metas estabelecidas. No compromisso relacionado diretamente à saúde, por exemplo, estabeleceu-se como objetivo a implementação de padrões e diretrizes para melhoria da interoperabilidade e usabilidade de sistemas relacionados ao Sistema Nacional de Vigilância Sanitária (CONTROLADORIA-GERAL DA UNIÃO, 2021).

Outra iniciativa é a GO (Global Open) Fair, que visa difundir e apoiar o desenvolvimento de atividades para adoção dos princípios FAIR (GO FAIR BRASIL, 2023). No Brasil, a iniciativa é coordenada pela GO FAIR Brasil, um dos escritórios regionais existentes pelo mundo. Existem diversas redes de implementação, entre elas a Rede GO FAIR Brasil Saúde, para coordenação das ações para implementação dos princípios FAIR no domínio da saúde (VEIGA et al., 2021).

As iniciativas relacionadas ao compartilhamento de dados aliadas a normas como a LGPD podem ser importantes mecanismos para a maior integração dos dados gerados por indivíduos e instituições e, com isso, potencializar os benefícios de *Big Data* para as diversas áreas do conhecimento.

3 TRABALHOS RELACIONADOS

Esse capítulo será dividido em quatro partes sendo, 1) Aplicação de ML na saúde e dengue, 2) Métodos de avaliação em saúde e dengue, 3) Características comuns dos trabalhos analisados e 4) Limitações e desafios.

Os trabalhos de aplicação de ML na área da saúde, em especial à dengue, evidenciam importantes fatores, como principais objetivos no uso de algoritmos nessa área, melhores algoritmos e fontes de dados mais utilizadas, assim como importante discussões sobre as limitações e desafios.

Em relação aos trabalhos relacionados à avaliação em saúde, procurou-se não se limitar à dengue, visto que muitos dos componentes avaliados são comuns à dengue e outras doenças, como o número de profissionais na atenção básica e número de unidades de saúde. Esses trabalhos descrevem as formas de coleta de dados e identificam os limites e desafios para aplicação de avaliações em saúde.

3.1 Aplicações de ML

A seguir, são apresentadas algumas aplicações de *machine learning* na área da saúde e, mais especificamente, no caso da dengue.

3.1.1 Área da Saúde

Assim como em outras áreas, a aplicação de ML na saúde pode trazer inúmeros ganhos. No entanto, essa área tem características que evidenciam ainda mais a importância do uso de algoritmos de ML. No tratamento de doenças, por exemplo, a precocidade do diagnóstico pode ter influência direta no desfecho, ou seja, quanto mais cedo a doença for identificada, maiores são as chances de cura do paciente.

Outro fator refere-se à diversidade dos dados normalmente utilizados na avaliação médica. Podem ser incluídos exames de imagens, dados laboratoriais, prontuários e, mais recentemente, dados advindos dos dispositivos *wearables*, como aqueles que podem coletar dados referentes aos hábitos dos pacientes. Dada a grande quantidade de informações sobre o paciente, como histórico, medicações, genoma,

atividades em redes sociais, o uso de algoritmos de ML pode ser útil como ferramentas de apoio a análise médica, embora sem garantia de “mágica”(BEAM; KOHANE, 2018).

A extração eficiente de informação desses dados pode auxiliar no correto diagnóstico de uma doença, assim como na prescrição adequada das intervenções necessárias para o tratamento dos pacientes. O uso de ML na área da saúde tem grande potencial para agregar valor na saúde e bem-estar das pessoas e cada vez mais pesquisas têm sido realizadas nessa área. Apenas considerando aplicações de *deep learning* e trabalhos de 2010 a 2015, nota-se um grande aumento de pesquisas usando essa forma de aprendizado de máquina em aplicações na saúde, em especial na análise de imagens médicas (RAVI et al., 2017).

Dado o seu potencial de processamento de grandes volumes de dados de fontes diversas, pesquisas com o uso de ML para diagnóstico de doenças têm aumentado. Uma revisão de literatura sobre ML para diagnóstico de mal de Parkinson identificou um 209 de estudos entre 2009 e 2019, sendo mais da metade publicada a partir de 2017. Dados dessa pesquisa apontam ganhos no uso de ML para diagnóstico mais fácil do mal de Parkinson no estágio inicial (MEI; DESROSIERS; FRASNELLI, 2021).

Em certos períodos durante a pandemia da COVID-19, o alto número de pessoas em atendimento médico gerou escassez dos insumos utilizados para realização de exames. Nesse cenário, muitas pessoas podem ter ficado sem o diagnóstico correto, causando subnotificação dos casos. A experiência mostrou que os países que obtiveram melhores resultados foram aqueles que conseguiram rastrear os casos de forma mais eficiente. Nesse sentido, pesquisas que desenvolvam métodos para correto diagnóstico podem ser muito úteis em cenários onde faltam recursos materiais e humanos, como uma pandemia. Uma pesquisa realizada em Israel desenvolveu um modelo para auxílio no diagnóstico usando apenas oito variáveis: sexo, idade ≥ 60 anos, contato com infectado e aparecimento (sim/não) de cinco sintomas (ZOABI; DERI-ROZOV; SHOMRON, 2021). Nessa pesquisa, o modelo foi treinado com dados de mais de 51 mil indivíduos, alcançando alta performance.

Outras aplicações envolvendo ML têm como objetivo o desenvolvimento de modelos de previsão. Na saúde, esse tipo de aplicação é fundamental para várias áreas,

como na adoção de intervenções para prevenção de doenças e planejamento para alocação de recursos, como preparação das unidades de saúde para uma epidemia.

Um estudo treinou um modelo a partir de dados de 1.029.366 pacientes para previsão de efeitos adversos causados pela diabetes (RAVAUT et al., 2021). Esse modelo utilizou o *Gradient Boosting Decision Trees* (GBDT) e foi testado com dados de 265.406 pacientes. Foram utilizados dados dos pacientes, como sexo e idade, além de dados relacionados à busca de atendimentos nas unidades de saúde. Para cada risco associado (por exemplo, hipoglicemia ou amputação), foi realizado um estudo da importância de cada característica utilizada no modelo. Além disso, os autores desenvolveram um estudo para avaliar os custos associados à ocorrência de cada efeito adverso. Este tipo de pesquisa é muito útil para demonstrar o impacto econômico que uma vigilância inadequada pode causar, além de identificar precocemente os riscos de efeitos adversos de modo que os pacientes tenham o acompanhamento adequado.

Com a maior disponibilidade de dados e o poder de processamento dos algoritmos sobre grandes massas de dados, pesquisas podem ser realizadas para gerar conhecimento sobre possíveis relacionamentos entre diversos tipos de variáveis e doenças. Um estudo utilizou dados demográficos, socioeconômicos e ambientais no desenvolvimento de um modelo para previsão de casos de COVID-19 (AHMAD et al., 2021). Os resultados mostraram a influência positiva da qualidade do ar sobre os casos, enquanto a média de idade e o IDH influenciaram negativamente. A descoberta dessas possíveis relações podem ser úteis para previsão de casos de doenças e consequente necessidade de planejamento para adequação dos serviços de saúde. Além disso, o fato de a qualidade do ar ser um fator preditor pode indicar mais um impacto negativo da poluição sobre a saúde humana.

Na área de saúde mental, algoritmos de aprendizado de máquina foram utilizados para prever o risco associado ao gênero para suicídios na Dinamarca (GRADUS et al., 2020). Este estudo realizou a análise de dados de mais de 14 mil casos entre 1995 e 2015, que encontraram resultados consistentes com o conhecimento sobre o assunto, mas com uma análise inédita de uma grande população. Ainda em saúde mental, um estudo teve como objetivo identificar sinais de depressão através de postagem de usuários no Twitter (VON SPERLING; LADEIRA, 2019). Com o uso de

Processamento de Linguagem Natural (NLP - *Natural Language Processing*) e aprendizado de máquina, os autores analisaram até 3200 mensagens de cada um dos 2941 usuários do estudo, divididos de forma balanceada nas duas classes (com ou sem depressão). Através de um aprendizado supervisionado, foram encontrados resultados condizentes com o encontrado na literatura.

Os estudos descritos anteriormente demonstram o ganho que a capacidade de processamento dos algoritmos pode agregar às atividades da área da saúde. O processamento de grandes volumes de dados, antes inviável devido às limitações de recursos, pode ser feito de forma a agregar velocidade ao diagnóstico de doenças assim como à vigilância em saúde. Além disso, o processamento com o uso de algoritmos de ML pode gerar *insights* que seriam inviáveis de serem realizados apenas com técnicas tradicionais de análise de dados.

3.1.2 Dengue

Assim como em outras doenças, a aplicação de algoritmos de ML no combate à dengue pode ter objetivos relacionados ao nível individual ou o coletivo. Os métodos utilizados para auxílio no diagnóstico podem ser muito úteis dentro do fluxo de atendimento na atenção em saúde em todos os seus níveis. Na atenção primária, por exemplo, a identificação de um potencial caso grave pode auxiliar no direcionamento do tratamento adequado para evitar a ocorrência de um caso grave. Além disso, a busca ativa por casos suspeitos pode se beneficiar de métodos automatizados, especialmente em cenários epidêmicos.

Pode-se dividir as aplicações de ML para dengue nas seguintes categorias:

- **Diagnóstico**, em que ML pode ser usada para triagem de pacientes
- **Previsão**, incluindo casos graves, epidemias e número de casos
- **Controle do vetor**, em que métodos de ML são usados em tarefas como identificação ou modelagem da população dos mosquitos
- **Identificação dos determinantes**, seja para fatores de risco de casos graves, incidência ou ocorrência de epidemias

- **Avaliação de intervenções**, onde mede-se a efetividade de uma intervenção ou um programa

Uma revisão sistemática identificou a prevalência dos modelos logísticos para diagnóstico de dengue, principalmente buscando avaliar os fatores associados ao risco de infecção e determinar os fatores para predição dos casos graves (HOYOS; AGUILAR; TORO, 2021). Em outro trabalho, foram empregados algoritmos de ML para prever o risco de severidade da dengue baseados nos dados de genoma de 102 pacientes brasileiros (DAVI et al., 2019). Como resultado de aplicação de uma rede neural, os autores alcançaram uma acurácia de 86% nos resultados, com sensibilidade de 98% e especificidade de 51% e, segundo conclusão dos autores, o método proposto poderia ser usado para prever risco da versão severa da dengue mesmo em pacientes não-infectados.

Com o uso de um método não-supervisionado, um estudo de 2019 teve como objetivo identificar características dos pacientes para criação de um critério de risco da severidade da dengue (MACEDO HAIR; FONSECA NOBRE; BRASIL, 2019). Através da análise de dados de 523 casos confirmados de dengue, quatro *clusters* foram identificados sendo a idade um critério chave para separação dos grupos. Segundo os resultados, o maior risco foi identificado no *cluster* composto principalmente por crianças de 5 a 15 anos.

Uma outra pesquisa aplicou diversos algoritmos de ML para criar um modelo preditivo para dengue (GUO et al., 2017). Essa pesquisa utilizou dados de dengue, informações meteorológicas e informações sobre buscas na Internet. Foram coletados dados de casos de dengue entre 2011 e 2014 da província de Guangdong, na China, que continham informações demográficas, histórico dos sintomas e sorotipo. Além desses, foram coletados dados de chuva, umidade e chuvas a partir de estações meteorológicas, bem como dados de buscas a partir de palavras-chaves associadas ao nome e sintomas da doença a partir do portal chinês Baidu. Dentre os algoritmos utilizados, foi identificado que o *Support Vector Regression* (SVR) alcançou os melhores resultados.

O mesmo algoritmo teve a melhor performance de predição em outro trabalho relacionado à dengue (SALIM et al., 2021). Neste estudo, com dados de 2013 a 2017

em cinco distritos em uma cidade da Malásia, foram utilizados como preditores variáveis climáticas (temperatura, umidade, velocidade do vento e chuva), além da semana do ano, o ano e o distrito. No modelo com melhor performance, o preditor de maior importância foi a semana do ano. Isso corrobora o fato da sazonalidade da doença, que tende a ocorrer nos mesmos períodos do ano. No entanto, essa variável pode perder sua força preditora nos cenários de mudanças climáticas devido à possibilidade de aumento e variação dos períodos com condições favoráveis para desenvolvimento do vetor.

Outra aplicação refere-se à modelagem de epidemias, em que os algoritmos são usados com objetivos de previsão de epidemias. Dada a possibilidade de identificação antecipada de uma epidemia, esse tipo de aplicação pode ser muito utilizado para auxílio no planejamento de ações e recursos dos programas de combate à dengue. Estudos nessa área podem considerar diferentes aspectos, como o tipo de análise (por exemplo: análise espaço-temporal da dengue), objetivos (distribuição do vetor, previsão de morbidade e mortalidade) e fontes de dados utilizadas (interações dos usuários via ferramentas de buscas e redes sociais) (HOYOS; AGUILAR; TORO, 2021).

O potencial do uso de ML para previsão de epidemias de dengue foi evidenciado em um estudo realizado com dados de 11 cidades de Bangladesh (DEY et al., 2022). Nesse estudo foram utilizadas variáveis climáticas e números de pacientes em dois modelos de regressão: SVR (*Support Vector Regression*) e MLR (*Multiple Linear Regression*), com melhor desempenho do primeiro modelo. Uma desvantagem apontada nesse estudo sobre esse modelo refere-se à lentidão em aplicações em larga escala, o que poderia ser uma barreira para sua adoção. Assim como em outros estudos no Brasil, esse estudo apontou a sazonalidade da doença, com maior ocorrência nos meses chuvosos e de maior temperatura.

Existe uma certa variabilidade nos algoritmos utilizados para previsões de taxa de incidência e epidemias de dengue. Uma revisão de literatura identificou em 33 trabalhos aplicações de diversos tipos de algoritmos, como redes neurais, árvores de decisão, regressão, séries temporais, entre outros (SYLVESTRE et al., 2022). A mesma variabilidade foi observada entre os modelos de melhor performance, com leve

predomínio dos modelos de regressão nas tarefas de previsão de taxas de incidência. Entre as variáveis preditoras, as mais significantes foram variáveis climáticas (chuva, umidade e temperatura). Nos trabalhos sobre diagnóstico da dengue, os dois melhores preditores foram os sintomas febre e mialgia. No entanto, os resultados evidenciaram um crescimento no uso de outras fontes de dados, como ferramentas de buscas e redes sociais.

Uma outra revisão de literatura realizou a análise de 120 estudos relacionados ao controle vetorial (JOSHI; MILLER, 2021). As abordagens identificadas nos trabalhos analisados foram classificadas em três grupos, geoespacial, visual e som, sendo a primeira mais utilizada nas tarefas de previsão e as outras duas para atividades de classificação e clusterização. Entre as técnicas de ML mais utilizadas destacaram-se árvores de decisão, redes neurais, Maxent e SVM. Para as atividades de previsão, foram identificados diversos tipos de dados, como imagens aéreas e dados de sensores, além de dados ambientais (por exemplo: clima, cobertura do solo) e socioeconômicos (pobreza). Foram identificados alguns trabalhos que utilizaram fontes de dados de texto, como redes sociais e ferramentas de buscas, que poderiam prover informações sobre a disseminação do mosquito. Técnicas de *deep learning* foram utilizadas para detecção de ovos, larvas e mosquitos na fase adulta. Os autores apontam que a detecção de fontes de água parada é uma das técnicas mais promissoras no controle vetorial, com destaque para a aplicação de redes convolucionais. Entre as oportunidades para uso de ML no controle vetorial, são destacados o uso de ciência cidadã e dados de *crowdsourcing*, desenvolvimento de modelos e conjuntos de dados abertos, além da adoção de novas técnicas como *transfer learning* e aprendizado por reforço.

3.2 Avaliação em saúde e Avaliação em saúde – dengue

A avaliação em saúde pode ser utilizada com diversos fins, como validar a eficácia de uma intervenção ou programa, verificar a adequação quanto às (ou quanto a) normas estabelecidas, monitorar indicadores em um processo de melhoria contínua, entre outras. Tão diversos quanto os objetivos são os mecanismos empregados para a execução de um processo de avaliação. Podem ser adotados questionários, entrevistas, observação e outros apoiados no método científico. Cada método tem suas vantagens e desvantagens, assim como maior adequação dependendo do objetivo e contexto.

Análise realizada na base de dados Scielo para trabalhos sobre avaliação em saúde entre 2017 e 2021 identificou que, dos 37 trabalhos selecionados, 16 (43,24%) utilizaram o questionário como forma de coleta de dados para avaliação, enquanto métodos baseados em entrevistas foram utilizados em 10 trabalhos (27,02%). Dada a aplicabilidade de cada método, o uso de métodos combinados é bastante comum, e os dados coletados por um método são avaliados em conjunto com dados oriundos de outro. No mesmo levantamento de literatura, foi identificado que cerca de 45,95% dos trabalhos utilizaram ao menos dois métodos para levantamento de dados.

Um estudo combinou dois métodos para avaliação da assistência pré-natal às gestantes de Joinville-SC (VAICHULONIS et al., 2021). Foram utilizadas entrevistas, para coleta de dados das pacientes, e revisão dos cartões das gestantes, documento utilizado para registro do atendimento das pacientes, como número de consultas, dados de imunização e testes. Foram estabelecidos seis indicadores de acordo com os critérios definidos no Programa de Humanização no Pré-natal e Nascimento (PHPN).

O uso de indicadores baseado nas normas estabelecidas é importante para verificar aspectos como a qualidade das ações implementadas. Além disso, dado que muitos critérios são estabelecidos com base em pesquisas científicas ou padrões internacionais, a avaliação desses critérios pode revelar possíveis problemas que coloquem em risco o bem-estar dos indivíduos. Por exemplo, o PHPN estabelece que a assistência pré-natal deve ser iniciada até o quarto mês da gestação. A avaliação realizada por VAICHULONIS et al. (2021) identificou o início tardio da assistência para cerca de 7,3% das mães entrevistadas, o que poderia ter impacto no acompanhamento das pacientes.

A mesma questão foi observada em outro estudo relativo ao mesmo programa, realizado com puérperas no Estado do Sergipe (MENDES et al., 2020). Nesse estudo, o índice de gestantes com início tardio foi ainda maior (43%) do que o estudo realizado por VAICHULONIS et al (2021). Através da análise de dados das gestantes, foram discutidas hipóteses acerca das possíveis relações desses resultados com o perfil das gestantes participantes do estudo (idade, escolaridade etc.).

O uso de critérios com base nas normas também pode servir para avaliação do grau de implantação de um programa. Uma pesquisa realizada para avaliação do

Programa Rede Mãe Paranaense (PRMP) definiu critérios de acordo com as metas estabelecidas pelo programa e, para cada critério, estabeleceu a relação entre indicador, meta e fórmula de avaliação (SANTOS et al., 2020). Nessa avaliação, os indicadores foram agrupados em quatro domínios e os resultados (índice avaliativo) foram calculados por domínio, além de um índice geral (proporção sobre a meta). A organização das ações em domínios pode ser muito útil para avaliar os pontos fortes e fracos de um programa, gerando informações importantes para a gestão do programa. Com essas informações, por exemplo, a gestão pública pode planejar a alocação de recursos para desenvolvimento dos pontos passíveis de melhoria do programa em um determinado domínio, além de monitorar o desempenho de cada domínio ao longo do tempo.

Quando utilizado o método de análise dos dados existentes em controles como documentos ou sistemas de informação, podem existir limitações quanto a problemas de subnotificação ou baixa completude das informações preenchidas. No estudo para avaliação da eficácia do Programa Rede Mãe Paranaense, foi identificado que, na maioria das vezes, o profissional não registrava as informações de forma adequada no cartão de saúde da criança (SANTOS et al., 2020). Outro estudo realizou a análise dos cartões de gestantes atendidas em um hospital universitário e cerca de 80% dos cartões tiveram o grau de completude classificado como ruim (COÊLHO et al., 2015). Além dos problemas no próprio atendimento que a falta de algumas informações pode causar, como prejuízos no acompanhamento das gestantes, a ausência de informações necessárias podem limitar a avaliação da intervenção e/ou programa de saúde.

Muitas avaliações utilizam os dados provenientes dos sistemas de informação governamentais, como o SINAN ou Sistema de Informações sobre Nascidos Vivos (SINASC). Fundamentais para acompanhamento da execução das políticas públicas, esses sistemas são utilizados para integrar os dados informados pelos entes federativos. Nesse processo de integração, são executadas atividades de validação dos dados antes da disponibilização nos sistemas públicos, como o TABNET (SILVA, 2009). Dado o caráter descentralizado do SUS, o monitoramento das políticas e a vigilância em saúde dependem do uso desses sistemas pelos entes federativos conforme preconizado, como os prazos para envio dos dados. Com base nisso, processos avaliativos podem ser conduzidos tendo como objeto o uso desses sistemas buscando a verificação de vários

aspectos, como qualidade dos dados e grau de adesão. Uma avaliação realizada sobre o Sistema de Informações do Programa Nacional de Imunizações (SI-PNI) adotou diversos critérios, como qualidade dos dados, simplicidade e aceitabilidade (SILVA et al., 2021). Entre os resultados encontrados, foram destacados problemas de completude e consistência nos dados, além da falta de adesão de muitos municípios ao sistema no período analisado (2017). Dessa forma, a avaliação de um programa pode ser limitada devido à baixa qualidade e cobertura dos dados. Avaliações quanto ao contexto, por exemplo, poderiam ser prejudicadas, especialmente em um país tão diverso como o Brasil.

Uma revisão de literatura realizou a análise de 10 estudos de avaliação do Programa Nacional de Alimentação Escolar (PNAE) (BICALHO et al., 2021). Nesse estudo, foi realizada a avaliação da qualidade dos indicadores utilizando o instrumento AIRE (*Appraisal of Indicators through Research and Evaluation*), em que 20 itens são agrupados em quatro domínios: Propósito, relevância e contexto organizacional; Envolvimento das partes interessadas; Evidência científica; e Evidência adicional, formulação e uso. Segundo resultados, embora os estudos tenham alcançado altos níveis de qualidade nos dois primeiros domínios da avaliação, o mesmo não foi observado nos outros dois, em que faltaram descrições suficientes das evidências formais que apoiavam o uso dos indicadores.

Além do embasamento na literatura, o modelo utilizado na avaliação pode ser validado por especialistas de domínio, grupo de profissionais ou pesquisadores com conhecimento sobre a área a ser avaliada. Uma pesquisa para avaliação sobre gestão do PNAE utilizou revisão de literatura e um painel de especialistas para validar o modelo avaliativo composto por cinco indicadores e nove medidas (PORRUA et al., 2020). Outro estudo utilizou-se da validação por especialistas dos instrumentos usados na avaliação do Programa de Enfrentamento às Doenças Negligenciadas para o controle da esquistossomose (DUBEUX et al., 2019). Nessa pesquisa, o programa foi avaliado quanto à duas dimensões do modelo de Donabedian (estrutura e processo), para as quais foram definidos 36 indicadores com base em documentos institucionais ou definidos em literatura. Ademais, o modelo de avaliação foi submetido à verificação de especialistas de modo a aumentar a validade do modelo.

A questão dos indicadores também é descrita em uma pesquisa para avaliação da rede de diagnóstico laboratorial do Programa de Prevenção e Controle das Hepatites Virais no estado de São Paulo (MARQUES; CARVALHEIRO, 2017). Nesta pesquisa, foram avaliados três componentes do programa (diagnóstico, vigilância e assistência) e, para cada componente, selecionados os procedimentos adotados e os indicadores relacionados. Ademais, foi verificada a correspondência entre os indicadores do componente diagnóstico em relação aos demais, considerando a relação entre processo e resultado, duas das três dimensões do modelo de Donabedian.

A avaliação dos serviços de saúde baseada na tríade estrutura, processo e resultado de Donabedian foi realizada em outro estudo para avaliação dos hospitais que participaram do Programa Nacional de Avaliação de Serviços de Saúde (PNASS 2015-2016), para as dimensões estrutura e processo do modelo (CHAVES et al., 2021). Nesse estudo foram definidos 30 critérios com base em cinco dimensões, onde cada critério era avaliado de forma binária (SIM/NÃO). Para comparação dos resultados, os hospitais foram caracterizados segundo quatro aspectos: tipo de gestão, esfera administrativa, nível de hierarquia e porte hospitalar. Esse tipo de classificação dos indivíduos avaliados permite avaliar os resultados quanto a diferentes aspectos. Por exemplo, no quesito esfera administrativa, foi identificado que a esfera pública federal alcançou a maior nota em relação aos demais (estadual, municipal e privada). Outra análise foi relacionada ao porte do hospital, com os melhores resultados sendo encontrados em hospitais de grande porte. A análise dos resultados da avaliação por grupo de indivíduos pode revelar informações importantes para a gestão, como identificar os fatores críticos que possibilitam melhor desempenho em um grupo em relação a outros.

A utilização das dimensões do modelo de Donabedian permite a análise do impacto das condições (estrutura) e ações (processo) nos resultados alcançados, da mesma forma que se pode avaliar as relações entre a estrutura e processos. Uma pesquisa sobre o uso da imunoglobulina palivizumabe para tratamento da infecção causada pelo vírus sincicial respiratório incluiu essas três dimensões na avaliação realizada no estado de São Paulo (GONÇALVES et al., 2018). Os resultados foram avaliados em 693 crianças participantes do programa quanto a ocorrência de hospitalização em UTI por doenças ou sintomas respiratórios. Para isso, foram

estabelecidas três hipóteses para verificar as relações entre falhas da aplicação (processo) e o desfecho (resultado). Embora tenham sido identificados pontos de melhoria na estrutura, essa não se mostrou fator limitante para os locais com processos avaliados como parcialmente adequados ou inadequados. Na mesma pesquisa foi identificado que os problemas nos processos foram fatores de risco para internações, em que cada falha nas aplicações aumenta 30% em média a chance de internação.

Outro estudo fez uso das três dimensões do modelo de Donabedian para avaliação do Programa de Controle da Hanseníase (PCH) em um município do estado de Pernambuco (SOUZA et al., 2017) . Esse estudo utilizou o tipo de análise de implantação para avaliar a relação entre o grau de implantação do programa sobre os efeitos observados. Para desenvolvimento dos critérios, foi desenvolvido um modelo lógico do programa para descrever a relação entre estrutura, processo e resultado para cada componente do programa (gestão, assistência à saúde, vigilância epidemiológica e educação em saúde e comunicação), bastante relacionados aos componentes do PNCD. Os dados foram coletados através de observação, entrevistas e do SINAN e os parâmetros para avaliação do grau de implantação foram coletados nas normas do programa. Foram identificadas deficiências nos indicadores relativos à estrutura, o que poderia gerar impactos negativos na assistência. Além disso, os problemas na vigilância e controle da doença podem estar associados aos altos valores dos indicadores de resultados. No entanto, dos quatro componentes avaliados, a vigilância obteve melhores resultados do que os demais, o que poderia ser utilizado pela gestão para desenvolvimento de ações para melhoria dos pontos fracos na implantação do programa, considerada “incipiente” pela pesquisa realizada.

A avaliação quanto ao grau de implantação também foi realizada em outra pesquisa, com o objetivo de avaliar os componentes das dimensões estrutura e processos do programa de controle da esquistossomose (COSTA et al., 2017). Assim como o trabalho de Souza et al. (2017), também foi elaborada nessa pesquisa um modelo lógico para descrever a relações entre a estrutura, processos e resultado para os componentes do programa. Alguns componentes desse programa também são comuns ao PNCD, como vigilância epidemiológica e educação e comunicação. A matriz de julgamento foi submetida a um comitê de especialistas, ponto considerado como importante para validação dos indicadores utilizados na avaliação (BICALHO

et al., 2021). Outra diferença nesse estudo foi a execução da avaliação em três municípios, enquanto o estudo de Souza et al. (2017) foi realizado com estudo de caso único. Os dados foram coletados através de um questionário estruturado aplicado a diversos profissionais de saúde das localidades avaliadas. Entre os resultados, foi identificado o baixo grau de implantação do componente educação em saúde, que também ocorreu no estudo sobre implantação do PCH (SOUZA et al., 2017). Esse componente, vale ressaltar, também está presente no PNCD. Assim, o baixo grau de implantação desse componente pode ter impacto em várias doenças, como a hanseníase, esquistossomose e a dengue.

Em relação à dengue, diversos estudos são conduzidos para avaliação das intervenções adotadas. Assim como em outras questões de saúde, as pesquisas podem ser desenvolvidas para avaliar um programa como um todo, uma intervenção ou grupo de intervenções ou ainda um componente específico. Um estudo realizou a avaliação do grau de implantação do PNCD em dois municípios do estado do Mato Grosso do Sul (COSTA; CUNHA; COSTA, 2018). Nesse estudo, foi desenvolvida uma matriz de análise e julgamento que contou com a validação de especialistas e continha 68 indicadores das dimensões estrutura e processo do modelo de Donabedian. Com os dados coletados através da aplicação de questionários junto a 383 profissionais de saúde, o PNCD foi considerado parcialmente implantado nos dois municípios. Através da análise dos resultados por componente do PNCD, diferenças nos resultados foram relacionadas a possíveis diferenças contextuais, como falta de recursos humanos e equipamentos para o desenvolvimento das ações.

Uma outra pesquisa realizou a análise do grau de implantação do componente Controle Vetorial no município de Porto Seguro (Bahia), na qual foram avaliados diversos critérios quanto às dimensões estrutura e processo (QUARESMA, 2017). Ademais, na avaliação, as variáveis relacionadas ao contexto foram consideradas, incluindo as condições sociodemográficas e saneamento básico. Assim como em outros estudos de avaliação de programas, foi desenvolvido pela autora o modelo lógico do componente controle vetorial. Em relação aos dados, as coletas foram realizadas através de entrevistas, observações de campo e documentos e relatórios oficiais. Embora as regiões do município tivessem suas particularidades, o estudo identificou que o programa foi implantado de forma homogênea no município.

Contudo, as características específicas das regiões demandam fluxos de trabalhos mais aderentes às suas realidades, como as regiões com limitações de acesso aos imóveis quanto ao tipo (residencial/comercial) e a questões relacionadas à segurança.

Uma revisão sistemática foi conduzida para comparar os efeitos de diferentes tipos de ações para controle vetorial (ERLANGER; KEISER; UTZINGER, 2008). Para comparação dos resultados, foram utilizados três índices para avaliação da infestação: o Índice de Breteau (IB), *House Index* e o *Container Index*. Além disso, para avaliação da efetividade dos programas, foi utilizada a incidência de casos como indicador. Num paralelo com o modelo de Donabedian, esses quatro indicadores estariam na dimensão *resultados*, enquanto as ações de controle vetorial analisadas estariam na dimensão *processos*.

De acordo com os resultados, as intervenções baseadas em controle ambiental mostraram-se menos efetivas do que aquelas baseadas em controle químico ou biológico. Há de se ressaltar, no entanto, que 9 dos 16 estudos analisados sobre controle químico foram realizados antes de 1983. A abordagem integrada, ou seja, a combinação de diversos tipos, mostrou-se mais eficaz do que as abordagens isoladas e, segundo os autores, poderiam ser mais sustentáveis devido ao fato de serem baseadas no engajamento da comunidade. Em relação aos impactos dos tipos quanto à incidência de casos, a quantidade de estudos não foi suficiente para o desenvolvimento de conclusões. Devido à ocorrência dos surtos em períodos irregulares, os autores sugerem que a incidência como medida de eficácia para as ações de controle vetorial pode ser mais significativa para avaliações em períodos longos.

Em revisão de literatura mais recente, dois indicadores foram adotados para avaliação dos resultados das ações de controle vetorial: índice de infestação e incidência de casos (BOWMAN; DONEGAN; MCCALL, 2016). Dos estudos analisados na revisão, a maioria utilizou apenas os índices vetoriais como forma de avaliação dos resultados. Esse fato representaria um problema nos estudos realizados de acordo com os autores, dada a baixa correlação desses índices com a transmissão da dengue. Para os autores, não há garantia de que as intervenções de controle vetorial se traduzirão em redução na transmissão. No entanto, foi apontado que a detecção de casos não é uma tarefa simples e nem de baixo custo.

Além disso, foi criticada a escassez de evidências para avaliação dos efeitos da pulverização, mesmo sendo recomendada pela Organização Mundial da Saúde para os surtos de dengue. O mesmo não foi observado para a avaliação das ações de visita domiciliar, para as quais foram encontradas evidências da eficácia nos estudos analisados. Ademais, foi destacada a necessidade de mais estudos considerando as ações baseadas em comunidade, como campanhas de limpeza de criadouros. Nos estudos analisados, essas ações foram avaliadas como um elemento dentro da avaliação de outras intervenções, não gerando evidências suficientes para avaliação da contribuição desse tipo de ação na redução dos vetores e transmissão da doença.

Nas avaliações das ações de combate à dengue, muitos fatores são apontados como razões para os resultados aquém das expectativas. No estudo desenvolvido por Eisen et al. (2009), são citados fatores como a urbanização descontrolada, o aumento da disseminação global e a falta de recursos adequados, que dificultam o controle sustentável do vetor da dengue (EISEN et al., 2009). Esse estudo propõe um modelo de avaliação para identificar métodos ineficazes e tecnologias ultrapassadas através de um ciclo de melhoria contínua. Nesse modelo, são defendidas as estratégias adotadas de acordo com o contexto local, já que as ações podem ter eficácia diferente dependendo do local em que são aplicadas.

A questão da continuidade das avaliações também é abordada em outro estudo, em que é citada a importância do monitoramento da susceptibilidade a inseticidas como forma de desenvolver conhecimento sobre os mecanismos de resistência (ZARA et al., 2016). Além disso, o monitoramento das situações regionais considerando outros fatores também pode ser importante para otimização dos recursos e adoção de medidas mais adequadas ao contexto.

3.3 Características Comuns na Avaliação em Saúde

De acordo com os trabalhos apresentados anteriormente, são observadas algumas características comuns a todas as atividades de avaliação, resumidas abaixo:

- Objeto: a avaliação pode ser realizada sobre uma intervenção ou um programa;

- Local: dado o custo de alguns mecanismos de coleta de dados, normalmente as avaliações são limitadas a uma área menor, como um hospital ou um município;
- Objetivo: uma avaliação pode ter um ou mais objetivos, como avaliação do grau de implantação ou avaliação da efetividade de um programa/intervenção. Nada impede, no entanto, que uma avaliação possa contemplar mais de um objetivo;
- Dimensões Donabedian: embora nem todas as avaliações descrevam claramente a utilização do modelo de Donabedian, as três dimensões do modelo podem ser identificadas, embora nem sempre com as três na mesma avaliação. É mais comum, no entanto, as avaliações utilizarem duas delas;
- Indicadores: o número de indicadores utilizados é bem variado, podendo ser relativos a metas e normas estabelecidas pelos programas ou definidos por quem está conduzindo a avaliação, com referência ou não por algum modelo ou base científica. Além disso, em diversos casos as matrizes de julgamento desenvolvidas para o estudo estiveram associadas a modelos lógicos dos programas, onde os componentes foram descritos quanto a recursos, atividades e indicadores;
- Métodos de coleta de dados: foi observado que os métodos adotados nas avaliações foram concentrados em três tipos, sendo questionário e entrevistas os dois mais comuns, embora a revisão documental seja bastante utilizada. Nesse terceiro tipo, destacam-se a coleta de dados em normas oficiais, documentos de saúde (cartões de vacinação, relatórios de unidades de saúde, etc.) e sistemas de informação (ex.: SINAN);
- Tipos de variáveis: as avaliações normalmente usam dados socioeconômicos e demográficos, para caracterização dos indivíduos, além de dados de saúde para verificação dos resultados da intervenção;
- Ferramentas para análise: nos trabalhos analisados identificou-se um predomínio de pacotes para análise estatística, como SPSS e R. Alguns

trabalhos mais recentes, no entanto, têm adotado técnicas de aprendizado de máquina;

- Estudo de caso: enquanto alguns estudos optaram pelo estudo de caso único, com avaliação em um município ou unidade de saúde, outros estudos adotaram análises usando casos múltiplos. Nesses casos, em geral foi realizada a caracterização de cada unidade de estudo.

3.4 Limitações e Desafios

Observou-se que a questão da qualidade dos dados foi apresentada pelos autores como limitação de boa parte dos estudos analisados. Em alguns casos, os dados fornecidos pelas pessoas entrevistadas podem representar limitação devido à múltiplas interpretações ou viés de memória (GARCIA et al., 2018; VAICHULONIS et al., 2021). Outros problemas foram relacionados à completude em sistemas de informação e documentos dos pacientes, como os cartões das gestantes. As razões para esse problema podem ser diversas, como limitações de usabilidade dos sistemas de informação, falta de capacitação ou disponibilidade dos profissionais responsáveis pela alimentação dos dados.

O fato é que nenhum modelo de avaliação está 100% imune a possíveis problemas nos dados, especialmente considerando-se a realidade brasileira, com administração descentralizada. Nesse cenário, é natural a adoção de controles próprios, geralmente desenvolvidos sem um protocolo comum. Além disso, os estados e municípios têm autonomia para criação de suas normas que, mesmo que sejam aderentes às nacionais, acabam refletindo no desenvolvimento de controles e dados geridos localmente. Assim, o alcance e qualidade dos processos de análise dos dados em escalas estaduais e nacional dependem necessariamente da forma como a integração dos dados dos entes federativos é feita.

Independente das limitações impostas pela qualidade dos dados, as avaliações em saúde podem revelar importantes informações para o monitoramento dos programas e intervenções. A condução dessas avaliações torna-se primordial para validação das ações adotadas além da identificação das razões para um evento de saúde, como surto de uma doença. Ademais, a frequência das avaliações em alguns

casos é ainda mais importante, como as doenças com frequência sazonal. Considerando a dinâmica em que as condições relacionadas mudam, como o clima e economia, a avaliação contínua dos programas pode ser útil para adaptação das ações frente às novas realidades apresentadas. As ações preconizadas no passado podem não ser mais adequadas no contexto atual e a identificação desse fato só é possível com a condução de avaliações de saúde.

No entanto, dependendo do método utilizado, o processo de avaliação contínua torna-se impraticável por limitações de recursos, visto que os custos de algumas abordagens são muito elevados, como condução de entrevistas a grandes grupos. Além disso, dada a velocidade com que as condições podem mudar e o fato de que alguns métodos demandam muito tempo para a sua execução, os resultados das avaliações podem levar demasiado tempo para disponibilização, análise e consequente tomada de decisão. Dessa forma, métodos alternativos são necessários como forma de incorporar aos métodos estabelecidos outros aspectos, como: potencial de integração de dados, velocidade na produção de informação e análise em larga escala dos efeitos das intervenções nos contextos em que são aplicadas.

4 DESENVOLVIMENTO DA PESQUISA

Este projeto de pesquisa foi dividido em três etapas (Figura 14), descritas nas subseções seguintes. A primeira etapa refere-se à definição do problema de pesquisa, para a qual foram realizadas atividades de revisão de literatura e interações com um comitê de especialistas. Em seguida foi desenvolvida a proposta de metodologia de avaliação do programa de controle de dengue com uma abordagem baseada em Ciência de Dados. Na terceira etapa, a proposta desenvolvida foi avaliada através de um estudo de caso para os municípios do estado de São Paulo.

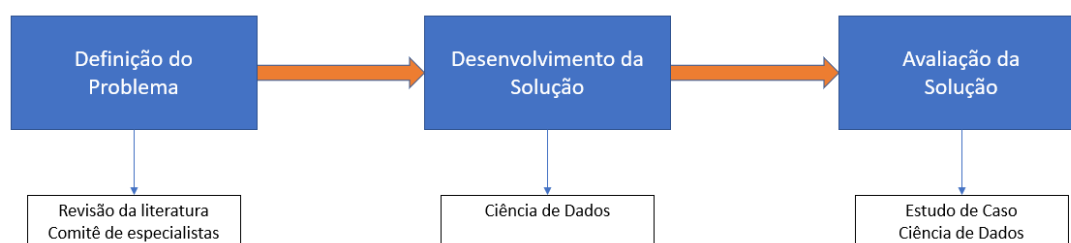


Figura 14 - Etapas da pesquisa e os métodos adotados em cada etapa.

4.1 Definição do problema

A definição do problema principal desta pesquisa foi realizada em duas fases: (1) estudo da literatura existente e (2) interação com especialistas de domínio. Essa divisão em fases foi necessária para se compreender, inicialmente, os conceitos relacionados ao domínio de aplicação desta pesquisa de doutorado, seguida pela análise dos processos existentes e desafios do programa de combate à dengue.

A primeira fase foi iniciada através da análise de publicações existentes sobre o campo de pesquisa Saúde Planetária. Essa análise tornou possível a compreensão dos aspectos relacionados a Saúde Planetária e permitiu identificar oportunidades de pesquisa. Como esse campo de pesquisa é bastante amplo e engloba diversos temas por ser uma abordagem sistêmica, foi conduzido um estudo de literatura sobre doenças causadas por vetores e, em especial, a dengue. Esse estudo teve como objetivo a compreensão dos aspectos relacionados à doença, como forma de transmissão, fatores determinantes e métodos de combate.

Embora projetos de Ciência de Dados possam ter a participação dos especialistas de domínio, a aquisição direta de conhecimento sobre o domínio pelo cientista de dados tem diversos benefícios, como analisar o problema sob outra perspectiva (diferente da área de formação do cientista), facilitar a interação com os especialistas e identificar as práticas usualmente adotadas no domínio, como o padrão ouro usado para determinado tema de saúde.

Em seguida, na segunda fase, foram realizadas diversas reuniões com especialistas do domínio com o objetivo de delimitação do problema de pesquisa, além de complementar o conhecimento adquirido na fase anterior. Essas reuniões foram realizadas com representantes de órgãos públicos da área da saúde das esferas municipal e estadual. No programa de combate à dengue, boa parte das ações do programa são realizadas pelos municípios, como aquelas relacionadas ao controle vetorial. Nesse sentido, conhecer a aplicação prática daquilo que é preconizado pelo programa permitiu identificar a dinâmica de implementação das ações de um município bem como suas principais dificuldades. Em nível estadual, responsável pela gestão e acompanhamento dos municípios, a interação com os especialistas de domínio foi necessária para identificar os principais desafios segundo a perspectiva dos gestores do programa, assim como explorar as bases de dados mantidas por esses órgãos. Ademais, a interação com representantes municipais e estaduais possibilitou o conhecimento acerca dos sistemas de informação utilizados pelas duas esferas governamentais no âmbito do programa de combate à dengue.

Além da interação com profissionais e gestores do programa de combate à dengue, foram realizadas reuniões com pesquisadores da área da saúde. Por meio dessas interações, foi possível ajustar aspectos do problema a ser pesquisado como, por exemplo, a escala espacial a ser analisada. Também foi possível identificar outras potenciais fontes de dados úteis para o desenvolvimento da pesquisa.

Dada a complexidade dos fatores relacionados à dengue, as interações realizadas com os especialistas de domínio contribuíram também para definição do problema de pesquisa segundo dois aspectos: 1) detalhamento dos objetivos específicos e 2) identificação de um problema de relevância científica e cujos resultados tenham menor tempo de implementação na gestão da saúde pública. Numa

perspectiva de pesquisas orientadas a problemas, a participação dos especialistas de domínio torna-se ainda mais essencial para incorporação dos resultados das pesquisas nas atividades da gestão pública, especialmente considerando-se que o fator tempo é primordial para o sucesso das ações relacionadas à saúde.

Nesse sentido, dentre as várias questões relacionadas ao programa de combate à dengue, foi definido que esta pesquisa de doutorado realizaria a avaliação do programa sob a perspectiva *Big Data*, ou seja, buscar responder às principais questões definidas em conjunto com os especialistas de domínio através da adoção de técnicas de Ciência de Dados para análise dos diversos conjuntos de dados históricos existentes. Com isso, foi definida a seguinte questão inicial de pesquisa: *Por que o programa de combate à dengue não alcança o êxito esperado?* Responder a essa questão pode trazer importantes informações para aprimorar a gestão e execução das ações do programa. No entanto, dada a complexidade dos fatores relacionados à dengue, provavelmente não exista uma resposta única para essa questão e a avaliação do programa pode compreender múltiplas perspectivas para identificação das razões do insucesso do programa. Ademais, a questão inicial descrita acima pode ser desmembrada em questões específicas, que podem ser definidas no início da pesquisa ou durante a condução das análises.

Em uma perspectiva normativa, por exemplo, pode-se identificar relações da eficácia com o grau de implantação das ações conforme normas definidas. As razões pelas quais as ações preconizadas não são plenamente aplicadas podem ser alvo de avaliações investigativas. Esse mesmo tipo de avaliação pode ser adotado para identificar possíveis impactos do contexto nos resultados considerando diversos aspectos, como socioeconômicos, climáticos, gerenciais etc.

No entanto, as avaliações em saúde comumente utilizam-se de instrumentos como questionários e observações de campo que, a despeito do seu valor, têm o custo e tempo de execução como fatores limitantes para aplicação em larga escala e produção de resultados rápidos. Ademais, o contexto de *Big Data* demanda abordagens alternativas para condução das avaliações em saúde de modo que possam agregar velocidade nas análises e poder de integração das informações geradas nos diversos sistemas relacionados aos serviços de saúde.

4.2 Desenvolvimento da solução: a metodologia de avaliação

A partir da questão identificada na etapa anterior, foi definida a proposta a ser desenvolvida nesta pesquisa: criação de uma metodologia de avaliação do programa de combate à dengue no cenário de *Big Data*. Para o desenvolvimento dessa proposta, foi adotada uma abordagem baseada em Ciência de Dados, composta por três fases, ilustradas na Figura 15. A primeira fase teve por objetivo a identificação das fontes de dados relacionadas à dengue e foi caracterizada pela grande interação com os especialistas de domínio, além de revisão documental das normas do programa, trabalhos relacionados e sistemas de informação oficiais.

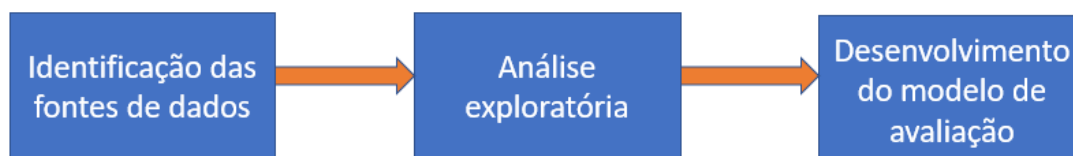


Figura 15 - Fases adotadas para o desenvolvimento da solução.

Em seguida, foi realizada uma análise exploratória das fontes de dados de forma a identificar os seguintes aspectos:

- Categoria dos dados;
- Mecanismos de acesso aos dados;
- Formas de disponibilização;
- Formatos dos dados;
- Escala espacial disponibilizada (estado, município etc.);
- Escala temporal (ano, mês, semana epidemiológica, dia);
- Intervalos de datas disponíveis.

Ao final dessa análise foi construída a documentação das fontes de dados relacionadas ao programa de combate à dengue para utilização nas fases seguintes da pesquisa.

Na terceira fase, que contou com a participação de um comitê de especialistas de domínio, foi desenvolvida a arquitetura da solução proposta para realização de avaliações em saúde. Essa arquitetura é composta de três partes, conforme Figura 16. A primeira parte consiste na definição do objetivo geral, como avaliação de um programa (ex.: combate à dengue). Na segunda parte, é construído um repositório comum de dados fundamentado nos conceitos, tecnologias e técnicas relacionadas a *Big Data*. Esse repositório pode ser alimentado por dados oriundos de sistemas transacionais, arquivos, dados de localização, redes sociais, entre outros.

Embora nada impeça o uso de múltiplas estruturas de dados, a adoção de um repositório comum para avaliações de uma mesma doença pode trazer múltiplos benefícios, como agilidade na condução das avaliações e diminuição do risco de ocorrência dos *data silos*, em que conjuntos de dados de diversas aplicações estão isolados uns dos outros. A ocorrência de *data silos* diminui as possibilidades de compartilhamento e colaboração de resultados entre pesquisas (PATEL, 2019).

Na segunda parte, *Avaliações*, múltiplas avaliações podem ser realizadas com conjuntos de dados coletados a partir do repositório de dados comum. As avaliações podem ter foco em um ponto específico de um programa, como uma ação de controle vetorial, ou ser realizadas para avaliação de todos os aspectos de um programa. Essa definição depende basicamente do alinhamento junto aos especialistas de domínio e demais interessados no objeto a ser avaliado.

Em algumas avaliações, parte dos dados pode ser coletada de outras fontes externas, como a aplicação de questionários junto a usuários dos sistemas de saúde. Nesse caso, esses dados coletados externamente podem alimentar o repositório comum, de forma que possam ser reutilizados em novas avaliações.

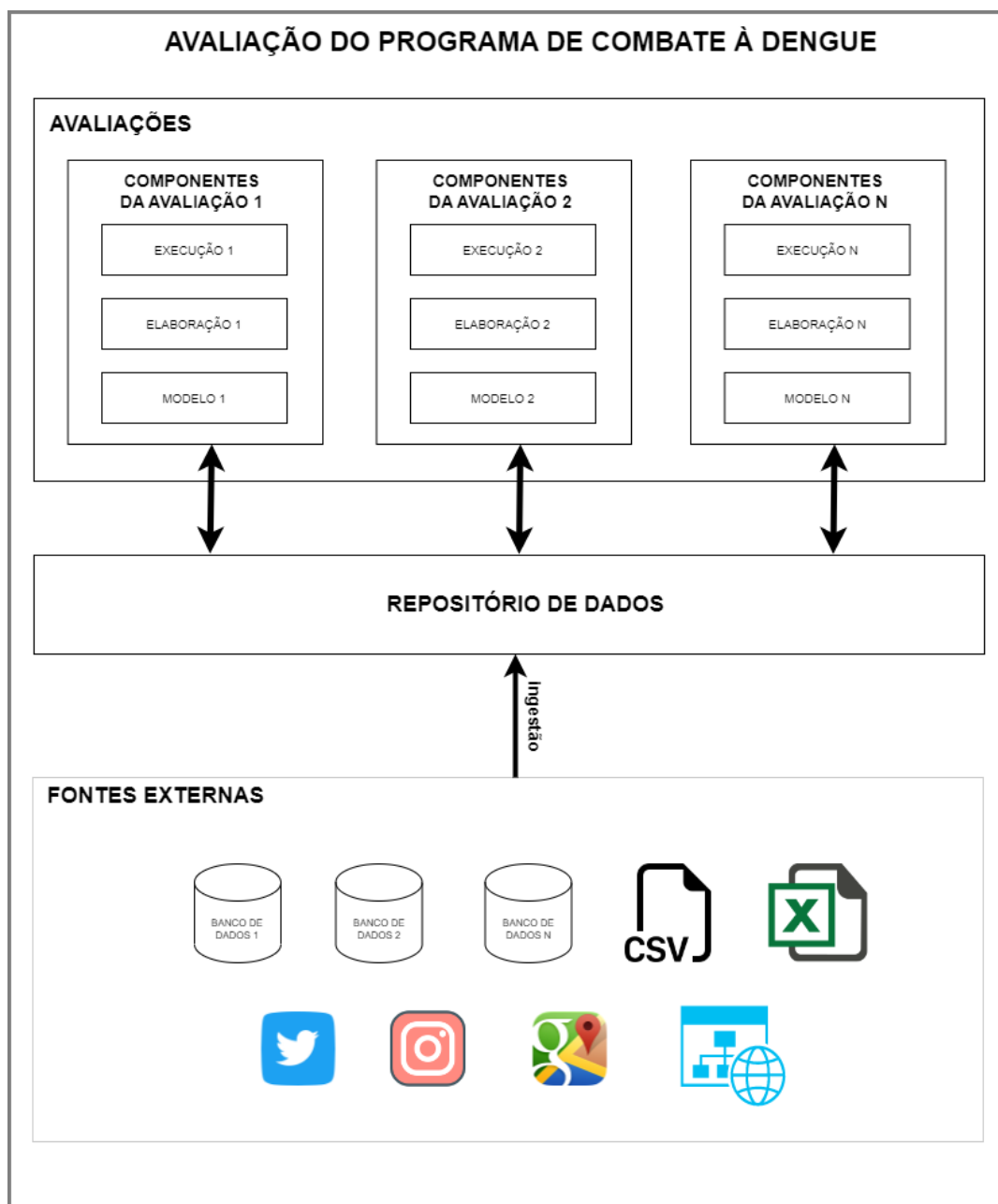


Figura 16 – Elementos da solução proposta nesta pesquisa.

Cada avaliação contém atividades específicas que, nesta proposta, foram agrupadas em três componentes: modelo, elaboração e execução. A partir do modelo de avaliação definido, os mecanismos para coleta de dados e análise são elaborados para posterior execução da avaliação e produção dos resultados. A seguir, são detalhados os aspectos relacionados ao repositório de dados e aos componentes da avaliação.

4.2.1 Repositório de dados

Como o próprio nome sugere, no componente Repositório de Dados, estão os dados necessários para desenvolvimento dos processos de avaliação. No caso do programa de combate à dengue, uma vez que são múltiplos os fatores relacionados a ela e existem diversos objetivos nas avaliações em saúde, esse repositório pode conter conjuntos de dados de diversos domínios, como epidemiologia, clima ou economia. A definição dos conjuntos de dados necessários para avaliação de uma doença é definida em conjunto com os especialistas de domínio, além das evidências encontradas na literatura. Nada impede, no entanto, que outros conjuntos de dados possam ser adicionados caso o objetivo da avaliação seja identificar possíveis novas relações de um fator com a dengue.

A construção do repositório base para realização de avaliações deve seguir os seguintes passos:

- Identificação das fontes de dados;
- Definição da arquitetura;
- Criação do repositório;
- Ingestão dos dados;
- Criação dos mecanismos de acesso.

Essas etapas são realizadas sequencialmente e em conjunto com os especialistas do domínio de interesse, no caso deste trabalho, a dengue.

4.2.1.1 Identificação das Fontes de Dados

Inicialmente, as fontes de dados relacionadas ao domínio de interesse devem ser identificadas, com o auxílio dos especialistas de domínio e de acordo com evidências encontradas na literatura. A documentação dessas fontes pode ser realizada de muitas formas, mas sugere-se, no mínimo, a documentação de algumas informações básicas:

- Nome da fonte;

- Domínio;
- Formas de acesso;
- Abrangência geográfica;
- Período disponível dos dados;
- Dados sensíveis?

Esses dados podem servir como base para definição da arquitetura, assim como para coleta e ingestão dos dados. A documentação deve ser atualizada constantemente, visto que os meios de acesso podem mudar, além da inclusão de novas fontes de dados de interesse. Além disso, a documentação deve considerar os princípios FAIR, indicando todas as informações necessárias para facilitar a reutilização dos dados.

4.2.1.2 Definição da Arquitetura do Repositório

Uma vez identificadas as fontes iniciais de dados, a atividade seguinte consiste na definição da arquitetura do repositório de dados. Essa etapa é fundamental para que a estrutura de dados necessária atenda aos requisitos inerentes ao domínio, assim como outros mais gerais, como segurança e facilidade de acesso. Ademais, uma arquitetura deve considerar a dinâmica com que os cenários podem mudar. Assim, prováveis mudanças não devem implicar em alterações drásticas no repositório se a arquitetura for definida adequadamente.

Para a construção deste repositório de dados, sugere-se a adoção da abordagem de um *Data Lake*, ou seja, são armazenados os dados brutos sem qualquer transformação. Como um mesmo conjunto de dados pode servir a diferentes propósitos, a manutenção dos dados em seus formatos originais transfere a decisão de quais transformações aplicar para cada avaliação.

A definição de qual arquitetura utilizar para construção do repositório de dados dependerá de muitos fatores, como aqueles relacionados às políticas das instituições. Mas algumas características devem ser comuns, como recursos eficientes para uso dos dados, flexibilidade e segurança das informações.

4.2.1.3 Construção do Repositório

A arquitetura definida no passo anterior servirá como base para criação do repositório, que pode estar armazenado em uma infraestrutura local ou em nuvem. A definição do local depende também do contexto, ou seja, das normas e políticas da instituição responsável pela elaboração das análises. Embora soluções em nuvem possam tornar mais fácil o acesso por usuários de diferentes localidades, a sua adoção depende da adequação às políticas das instituições. Dessa forma, a “melhor” solução é aquela que for mais aderente às características do contexto em que é aplicada.

4.2.1.4 Ingestão dos Dados

Uma vez criada a estrutura, procede-se com a alimentação do repositório a partir das fontes de dados definidas na primeira etapa. Esse processo, chamado de ingestão de dados, pode conter diversas atividades e a forma com que é realizado depende das características dos dados. A ingestão de dados pode ser classificada quanto ao mecanismo de operação (automática ou manual) e quanto à frequência (contínua ou em lote).

Em um modelo automático, são desenvolvidos *scripts* ou utilizadas ferramentas para coleta automática dos dados a partir das fontes. Considerando dados com fluxo contínuo de recebimento, como aqueles coletados de redes sociais ou sensores, a ingestão pode ser realizada assim que os dados estão disponíveis (*stream* ou contínua). Outra forma é realizar a ingestão em forma de lote, ou seja, faz-se a ingestão dos dados acumulados em um intervalo de tempo (por exemplo, a cada n horas). A forma manual de ingestão consiste no *upload* dos conjuntos de dados no repositório e, dependendo da arquitetura adotada, documentação manual das informações necessárias no repositório, como os metadados.

O repositório deve prover informações suficientes para descoberta de conjuntos de dados úteis para as análises, sem que seja necessária a análise detalhada do conteúdo dos conjuntos de dados. Nesse sentido, os itens do repositório podem ser descritos através de metadados considerando os princípios FAIR e as características das avaliações em saúde. Por exemplo, baseando-se no modelo Donabedian para avaliação

em saúde, poderiam ser adicionadas *tags* aos metadados para indicar a qual componente do modelo se refere determinado conjunto (estrutura, processo e resultado). Considerando ainda que as avaliações em saúde podem levar em conta o contexto em que as ações são aplicadas, uma segunda classificação dos conjuntos de dados poderia conter *tags* relacionadas ao tipo de variável de contexto, como clima, demográfica ou socioeconômica.

Informações adicionais podem ser incluídas aos metadados, como a descrição dos atributos dos registros. Essa descrição pode conter desde a simples identificação do atributo até informações mais detalhadas, como domínio dos dados, número de dados faltantes, entre outras. Como avaliações em saúde normalmente se utilizam do aspecto temporal, informações de data poderiam ser adicionadas aos metadados, como o intervalo de tempo a que os dados se referem. A definição de quais metadados adicionar é realizada pelos especialistas de domínio e deve ser flexível, ou seja, novas classificações dos conjuntos de dados podem ser adicionadas sempre que necessário. O objetivo principal dessas classificações é permitir uma qualificação rápida dos conjuntos de dados por quem vai conduzir as avaliações.

Outro ponto fundamental é em relação a segurança dos dados, principalmente devido ao fato de que dados de saúde podem conter informações de pacientes. Nesse sentido, mecanismos para proteção da privacidade e segurança devem ser adotados, como anonimização dos dados, controle de acesso e auditoria.

4.2.1.5 Criação dos Mecanismos de Acesso

Por fim, o repositório deve disponibilizar mecanismos de busca, descoberta e seleção de conjuntos de dados, de forma que os usuários possam executar consultas baseadas em filtros ao repositório, como os campos presentes nos metadados. Esses mecanismos devem considerar as restrições quanto a segurança dos conjuntos dos dados, como verificação se um usuário tem permissão para acesso a determinado conjunto de dados, mesmo para consulta.

Entre os exemplos de mecanismos que podem ser desenvolvidos estão incluídos portais WEB e as *Application Programming Interfaces* (API). Novamente, a

definição dos mecanismos a serem criados dependerá das regras de negócio da instituição.

4.2.2 Componentes da Avaliação

Uma avaliação é composta de três componentes, sendo 1) o **modelo** com a definição dos indicadores, 2) a **elaboração**, que consiste no planejamento das ações para coleta dos dados necessários, e 3) em que ocorre a **execução** da avaliação e geração dos resultados.

4.2.2.1 Modelo

O primeiro componente da avaliação, *Modelo*, refere-se ao modelo de avaliação, que contém um conjunto de critérios para avaliar um ou mais aspectos relacionados à doença. Além disso, os objetivos da avaliação devem ser definidos para delimitação do escopo das atividades a serem executadas, assim como identificação das fontes de dados necessárias. Em um processo orientado a problemas, os objetivos podem consistir em questões a serem respondidas para tomada de decisão (por exemplo, definição de área para investimento) ou investigação de um fato (por exemplo, que municípios não cumpriram com as normas?). Ademais, esse instrumento deve estar amparado em evidências científicas e/ou ser validado por comitê de especialistas para garantia da qualidade do processo avaliativo.

Assim, um modelo de avaliação pode conter os seguintes elementos:

- Objetivo: descrição clara do objetivo da avaliação;
- Hipótese: hipóteses ou questões que deverão ser investigadas;
- Alvo: qual programa ou intervenção está sendo avaliado;
- Categoria: o tipo da avaliação que será conduzida (p. ex: normativa ou avaliativa);
- Período: intervalo em que um programa/intervenção será avaliado;
- Forma: se será realizada de maneira contínua ou pontual;

- Dados: quais conjuntos de dados serão necessários para realização da avaliação. Essa descrição pode conter informações quanto a localidades, tipos de variáveis e quaisquer especificações cabíveis;
- Critérios: a matriz de julgamento utilizada na avaliação, podendo conter a categorização dos critérios quanto às dimensões do modelo de Donabedian ou outro tipo de categorias, como componente do programa. Ademais, para cada critério, deve ser indicada a base utilizada para sua seleção (norma, evidência científica e/ou comitê de especialistas).

Outras informações podem ser adicionadas ao modelo, como ferramentas e técnicas utilizadas para análise dos dados, formas de disponibilização dos resultados, documentos de comitês de ética, e outras que forem necessárias para devida documentação da avaliação.

4.2.2.2 Elaboração

Esse componente contém as atividades para desenvolvimento dos instrumentos para coleta dos dados. Caso seja necessária coleta de dados junto a usuários, instrumentos como questionários ou roteiros de entrevistas podem ser desenvolvidos. No caso de acesso aos dados do repositório, podem ser desenvolvidos meios para coleta dos conjuntos de dados necessários, que podem incluir *scripts*, consultas SQL.

Caso seja necessário, também podem ser desenvolvidos os meios para transformação dos dados de acordo com os meios de análise escolhidos. Nesse caso, são múltiplas possibilidades de transformação, como preenchimento de dados faltantes, normalização, *encoding* etc.

Destaca-se ainda que as avaliações podem ser conduzidas de forma pontual ou contínua e os tipos podem influenciar no desenvolvimento dos mecanismos de coleta e gestão dos dados. Para as avaliações realizadas de forma contínua, podem ser desenvolvidas estruturas analíticas, como os *Data Warehouses*, através de processos ETL que integrem em um banco de dados os dados selecionados do repositório. No caso de avaliações de forma pontual, podem ser desenvolvidos *scripts* para coleta dos dados ou mesmo acesso manual através das interfaces de acesso desenvolvidas no repositório de dados.

4.2.2.3 Execução

O componente *Execução* contém as atividades para execução da avaliação conforme os elementos definidos nos outros componentes. Para isso, devem ser desenvolvidas ou utilizadas as ferramentas de análises de acordo com o modelo proposto. Os métodos de análise adotados dependerão dos dados e objetivos utilizados na avaliação, podendo ainda ser uma combinação de diversos métodos, como aprendizado de máquina, estatística, entre outros.

A atividade final desse componente é a produção dos resultados da avaliação, que podem ser em forma de relatórios ou gráficos, mapas, usando outras técnicas/ferramentas de visualização computacional.

4.3 Avaliação da Solução

A terceira etapa desta pesquisa (vide figura 14) consistiu na Avaliação da Solução proposta desenvolvida. Essa avaliação se deu através de um estudo de caso, composto de dois experimentos, tendo como unidades de análise os 645 municípios do estado de São Paulo. A escolha desse estado é justificada pelo fato de ser o mais populoso do Brasil e com o maior número absoluto de casos de dengue no país. Além disso, esta pesquisa conta com parceria de pesquisadores ligados à Secretaria Estadual de Saúde que, além da validação dos instrumentos de avaliação do programa, possibilitou acesso aos dados das ações executadas pelos municípios do estado no âmbito do programa de combate à dengue. Esses pesquisadores formaram o comitê de especialistas de domínio que prestou suporte nas diversas atividades da pesquisa.

Os dois experimentos foram desenvolvidos de acordo com o ciclo de ciência de dados descrito na Figura 17 e baseado no método proposto por Shcherbakov et al. (2014), com etapas bem definidas, interação com especialistas de domínio e possibilidade de múltiplas execuções. O ciclo base foi adaptado para ressaltar o aspecto de geração de produtos e, com isso, deixar enfatizada a importância do desenvolvimento de pesquisas orientadas a problemas e na diminuição do tempo de implementação dos resultados no programa de combate à dengue.

A etapa inicial deste ciclo (passo 0) consiste na definição do objetivo principal do experimento, que é definido em conjunto com os especialistas de domínio. Esse objetivo servirá como ponto de verificação de cada rodada do ciclo e, enquanto não for satisfatoriamente alcançado, novas rodadas do ciclo poderão ser conduzidas. Além disso, devem ser definidos previamente os produtos resultantes da pesquisa, que podem ser em forma de relatórios, mapas e/ou *scripts* para automação de análises estatísticas e/ou modelos de aprendizado de máquina.

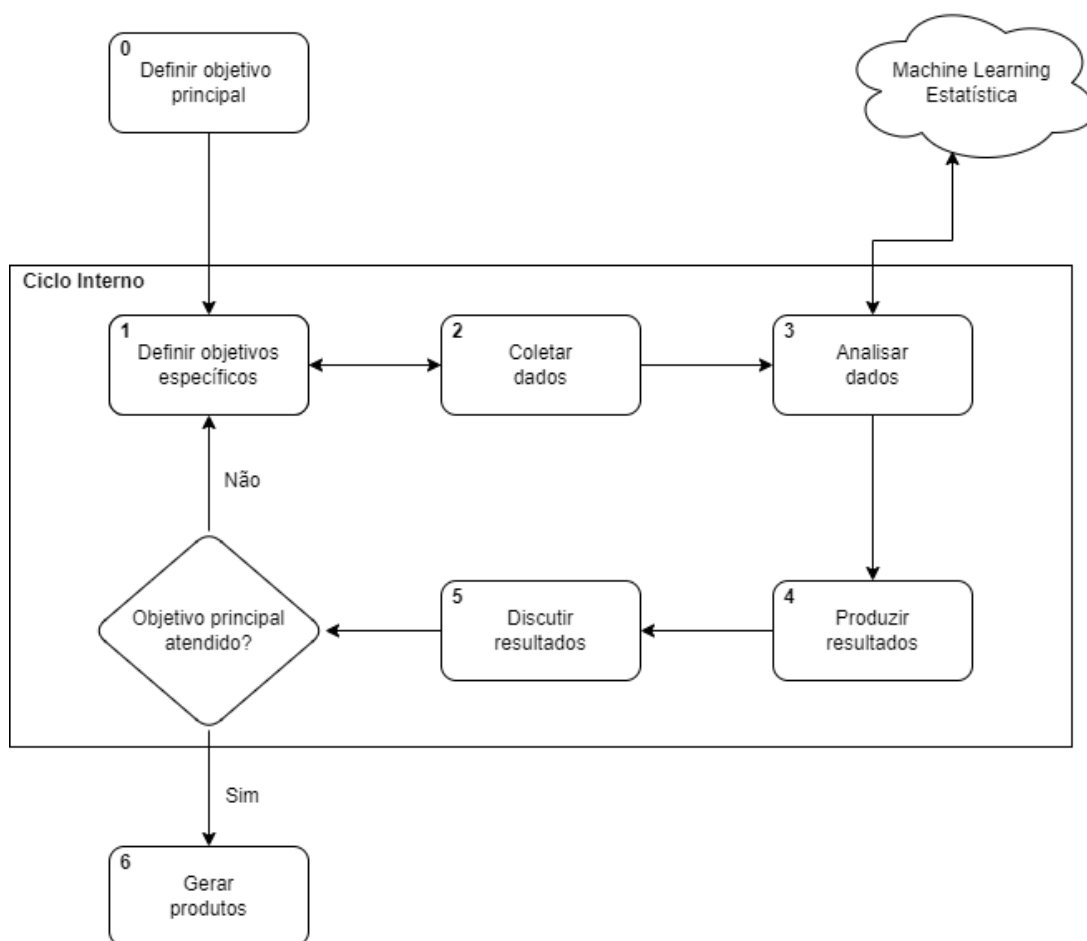


Figura 17 - Ciclo de Ciência de Dados usado no desenvolvimento dos experimentos.

Fonte: SHCHERBAKOV *et al.*(2014, adaptação nossa)

A partir do entendimento do contexto do domínio e definição do objetivo principal, objetivos específicos podem ser definidos. Para cada um desses, estudos específicos podem ser conduzidos, com execução dos passos 1 a 5 em um ciclo interno, conforme apresentado na Figura 17. Esses objetivos específicos podem ser definidos no início do experimento ou ao final de cada rodada do ciclo interno. Em uma pesquisa orientada a dados, novas questões podem ser definidas à medida que os dados existentes são analisados.

Em conjunto com a definição do objetivo específico no passo 1, as fontes de dados necessárias para condução deste estudo devem ser definidas. Em seguida, no Passo 2, os dados necessários para o estudo devem ser coletados a partir do repositório de dados criado para esta pesquisa. Em caso da ausência dos conjuntos de dados necessários no repositório, métodos de coleta de outras fontes deverão ser desenvolvidos para posterior ingestão dos dados coletados no repositório. Múltiplas formas podem ser adotadas para coleta de dados das fontes originais, como acesso a arquivos disponibilizados publicamente ou através de métodos automatizados, como acesso a uma *Application Programming Interface* (API). Nesse sentido, ressalta-se a importância dos conhecimentos em métodos computacionais de projetos de Ciência de Dados pois, além de desenvolver os algoritmos para as análises em si, o cientista de dados pode fazer uso de métodos alternativos para acesso aos dados quando esses não estiverem facilmente disponíveis, como desenvolvimento de *bots* para *web scrapping*, acesso a APIs etc.

O passo 3 do ciclo interno consiste na análise dos dados, incluindo-se atividades de transformação dos dados e execução de métodos para extração de informação. Esses métodos podem incluir, em conjunto ou isoladamente, a execução de algoritmos de aprendizado de máquina ou técnicas estatísticas. Não se considera adequado, nessa abordagem proposta, a definição de uma técnica única para a análise dos dados. A adoção de uma ou mais técnicas vai depender do objetivo do estudo intermediário assim como das características dos dados coletados. Ademais, esse passo pode ser composto de múltiplas execuções, de modo a se atingir os objetivos e critérios de qualidade previamente definidos como, por exemplo, o erro absoluto médio. Deve-se observar que os dados permanecem nos seus formatos originais em um repositório,

cabendo às tarefas de transformação serem executadas quando os dados são usados e realizadas de acordo com os objetivos da análise.

No passo seguinte do ciclo interno (4), devem ser gerados os resultados em formato que seja adequado para análise pelos especialistas do domínio. Assim como na definição da técnica utilizada no passo anterior, o objetivo do estudo e as características dos dados também podem ser fatores determinantes da forma de apresentação dos resultados. Por exemplo, estudos com dados que contenham elementos de localização poderiam ter seus resultados apresentados na forma de visualização espacial em mapas. Outro fator a ser considerado são os perfis dos especialistas de domínio, o que pode fazer com que a forma de apresentar os resultados seja variável de um domínio para o outro.

No último passo do ciclo interno, os resultados são discutidos em conjunto com os especialistas de domínio. Nessa etapa do estudo, deve ser verificado se a questão principal definida no passo 0 foi respondida ou se são necessários novos estudos. No segundo caso, então uma nova rodada do ciclo interno é iniciada com a definição de um novo objetivo. Mas, caso o objetivo principal seja atendido de maneira satisfatória após a execução da rodada do ciclo interno, então este é encerrado e são gerados os produtos definidos no passo 0 do ciclo.

5 ESTUDO DE CASO

Neste capítulo é apresentado o estudo de caso realizado, composto de dois experimentos. Inicialmente, a área definida para execução do estudo de caso é caracterizada, incluindo aspectos demográficos, socioeconômicos, meio ambiente e a situação de saúde, quanto a infraestrutura, investimentos e ocorrência de casos de dengue. Em seguida, são descritos os dois experimentos definidos para avaliação da metodologia proposta nesta pesquisa.

5.1 Área de Estudo

A metodologia proposta nesta pesquisa foi avaliada através de um estudo de caso para avaliação do programa de dengue nos 645 municípios do estado de São Paulo (SP), cuja caracterização será feita nas subseções a seguir.

5.1.1 Características Gerais

De acordo com dados do IBGE, o estado tem uma população estimada de 46.649.132 habitantes (2021), o que representa cerca de 21,87% da população brasileira (IBGE, 2022c). Com uma área de 248.219,49 quilômetros quadrados, SP tem uma densidade demográfica estimada de 187,94 hab/km² (2021).

Em relação a dados socioeconômicos, dados do Censo de 2010 indicam que o Índice de Desenvolvimento Humano (IDH) do estado foi de 0,783, acima do índice nacional (0,724). Em relação ao PIB, o valor per capita para o estado de SP foi de R\$ 52.992,00 em 2019 (FUNDAÇÃO SEADE, 2022a), também acima do valor nacional, que foi de R\$ 35.161,70 no mesmo período (IBGE, 2022c).

Outro índice que pode ser usado para caracterizar as condições de vida da população é a mortalidade infantil, para a qual a taxa para SP em 2019 (10,93 por mil nascidos vivos) foi menor do que a taxa nacional no mesmo período (13,3 por mil nascidos vivos) e abaixo da taxa na Região Sudeste (11,9 por mil nascidos vivos) (FUNDAÇÃO SEADE, 2022b; SECRETARIA DE VIGILÂNCIA EM SAÚDE,

2021). Analisando-se os municípios do estado, nota-se grande variação nessa taxa, conforme Figura 18.

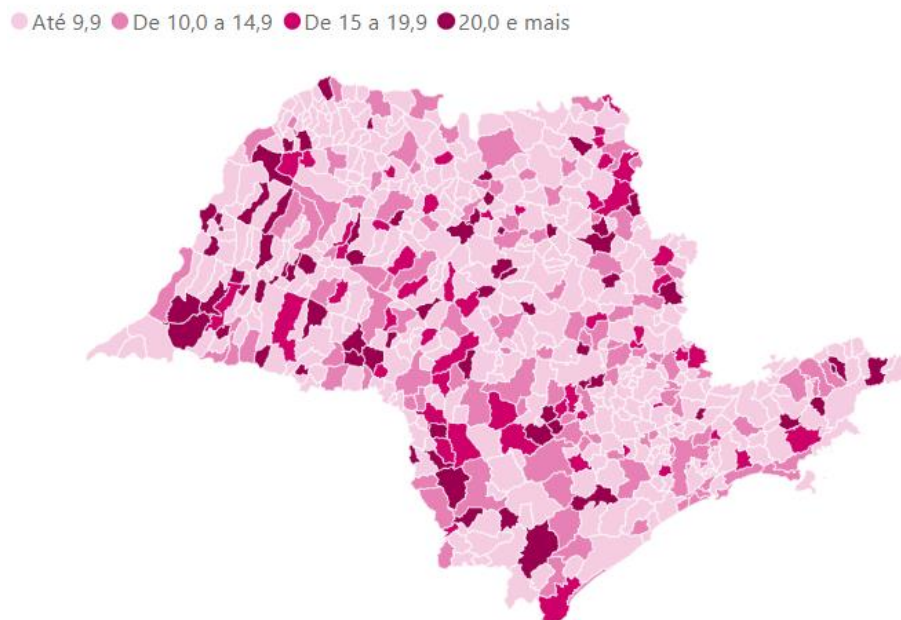
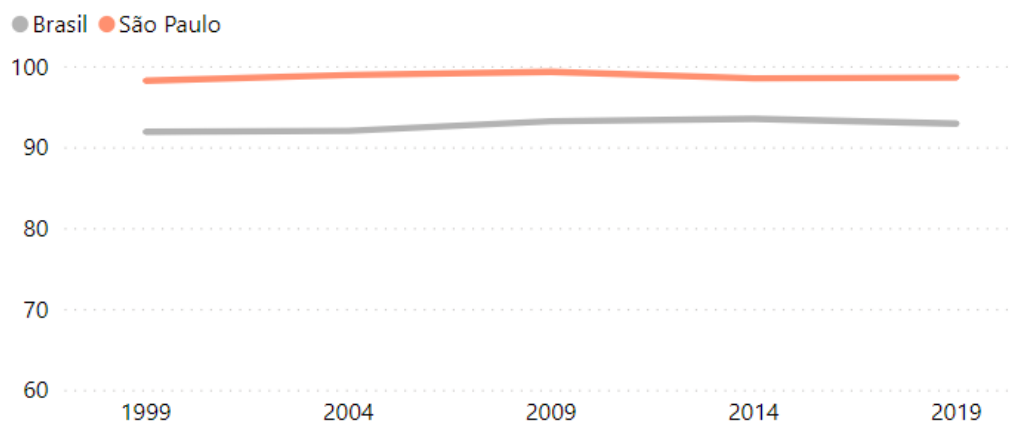


Figura 18 - Taxa de mortalidade infantil para os municípios do Estado de São Paulo em 2020.

Fonte: (FUNDAÇÃO SEADE, 2022b)

Em relação ao saneamento, cerca de 98,6% da população urbana do estado é atendida com rede de distribuição de água e em torno de 95% é atendida por rede de esgotamento sanitário, segundo dados de 2019 da Fundação Seade (FUNDAÇÃO SEADE, 2022c). Nos dois índices, os valores do estado de SP estão acima da média nacional, conforme Figura 19, que mostra as variações nos índices nos últimos 20 anos.

População urbana com rede de distribuição de água (%)



População urbana com rede de esgotamento sanitário (%)

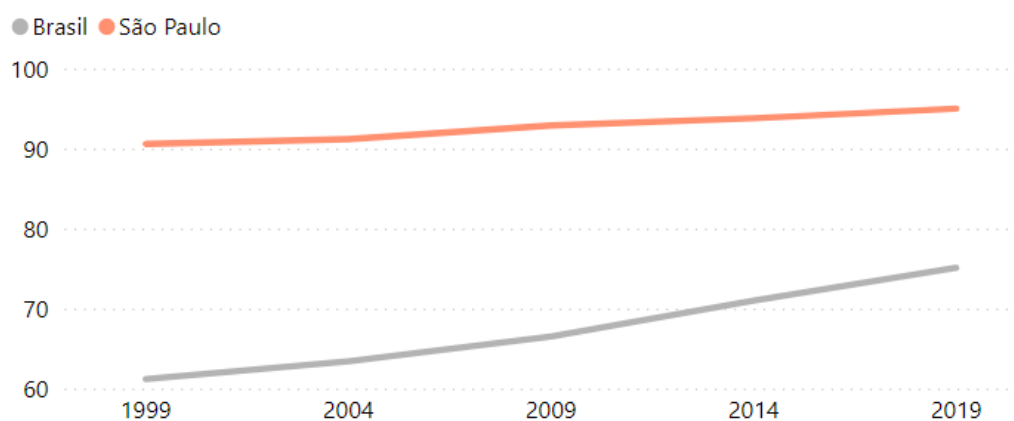


Figura 19 - Percentual da população atendida em saneamento entre os anos de 1999 e 2019 para o Estado de São Paulo e Brasil.

Fonte: (FUNDAÇÃO SEADE, 2022c).

Cerca de 81% dos municípios do estado têm grau de urbanização acima de 80%, segundo dados de 2022 (FUNDAÇÃO SEADE, 2022d). A Figura 20 mostra a variação do grau de urbanização dos municípios de SP, na qual é possível notar a predominância dos municípios mais urbanizados na faixa que vai da Região Metropolitana de São Paulo até a Região Nordeste do estado, coincidindo com uma das principais rodovias do estado: a Rodovia Anhanguera.

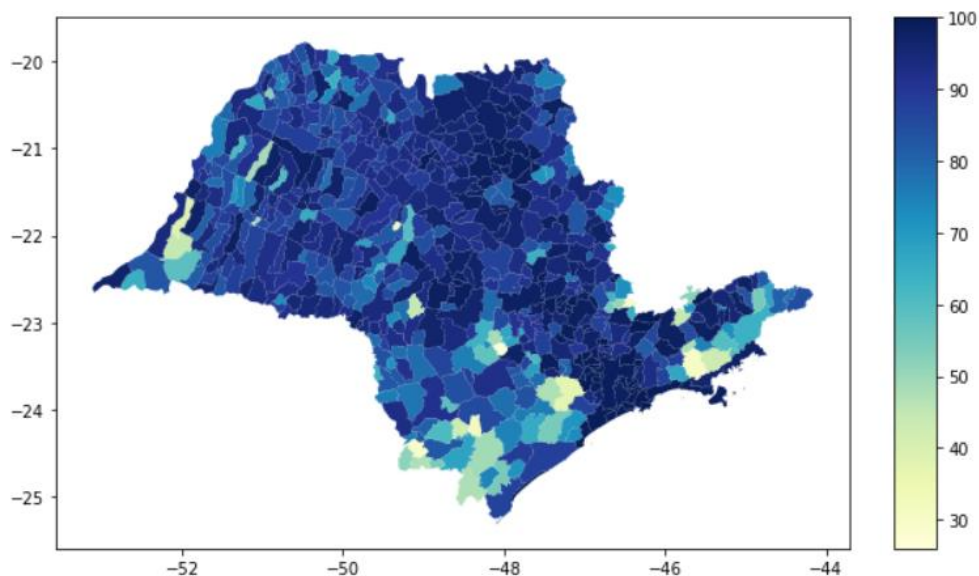


Figura 20 - Grau de Urbanização nos municípios do estado de São Paulo (2022).

Fonte: Elaborado pelo autor.

Em relação às condições de vida, foi desenvolvido um estudo para cálculo do Índice Paulista de Responsabilidade Social (IPRS), baseado nos mesmos critérios do Índice de Desenvolvimento Humano (IDH) (ASSEMBLÉIA LEGISLATIVA DO ESTADO DE SÃO PAULO, 2022). Por esse índice, os municípios do estado de São Paulo foram classificados em cinco grupos de acordo com os níveis de riqueza e indicadores sociais (Tabela 9).

Tabela 9 – Descrição dos grupos do Índice Paulista de Responsabilidade Social.

Grupo	Descrição
Desiguais	municípios com níveis de riqueza elevados, mas indicadores sociais insatisfatórios (longevidade e/ ou escolaridade baixo)
Dinâmicos	municípios com índice elevado de riqueza e bons níveis nos indicadores sociais (longevidade e escolaridade médio /alto)
Em transição	municípios com baixos níveis de riqueza e indicadores intermediários de longevidade e/ou escolaridade (níveis baixos)
Equitativos	municípios com baixos níveis de riqueza, mas bons indicadores sociais (longevidade e escolaridade médio/alto).
Vulneráveis	municípios mais desfavorecidos do Estado, tanto em riqueza como nos indicadores sociais (longevidade e escolaridade baixas)

Fonte: (ASSEMBLÉIA LEGISLATIVA DO ESTADO DE SÃO PAULO, 2022)

Conforme Figura 21, observa-se um padrão regional das condições dos municípios, no qual os municípios mais vulneráveis estão localizados principalmente nas regiões administrativas sul de Registro e São José dos Campos e aqueles com melhores condições (“Dinâmicos”) estão localizados principalmente nas regiões administrativas Central e de Campinas. Em relação a região metropolitana de São Paulo, onde se concentra boa parte da população do estado e cidade mais populosa do país, verifica-se uma concentração de municípios classificados como “Desiguais”.

Grupos IPRS ● Desiguais ● Dinâmicos ● Em Transição ● Equitativos ● Vulneráveis

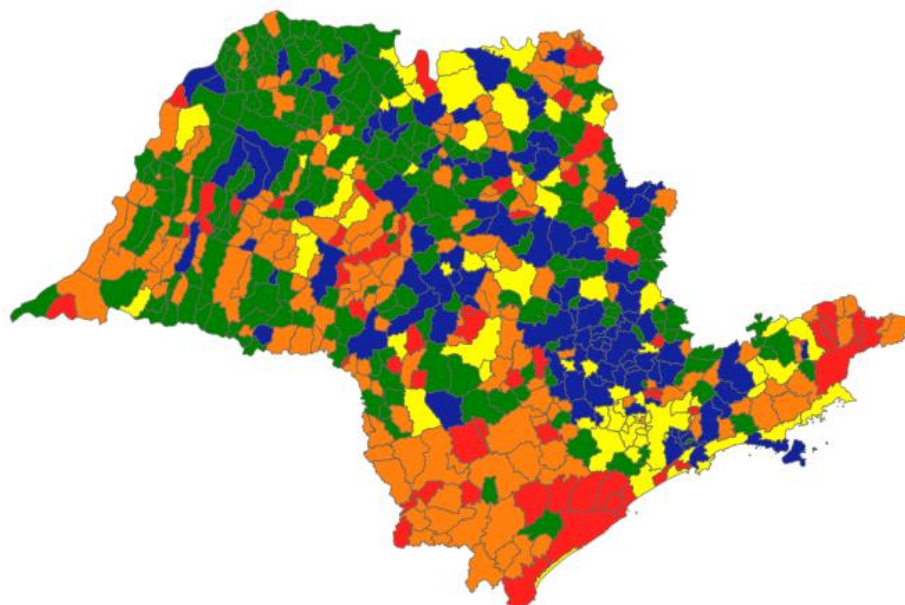


Figura 21 - IPRS dos municípios de SP em 2018.

Fonte: (ASSEMBLÉIA LEGISLATIVA DO ESTADO DE SÃO PAULO, 2022)

5.1.2 Clima

Usando a classificação de Köppen, o clima predominante no Norte e Noroeste do estado é o Aw (clima tropical com inverno seco) e no sudoeste do estado há a predominância do clima subtropical, com verão quente (Cfa, na classificação de

Köppen) (ROLIM et al., 2005). A temperatura média anual varia de 14,08 a 23,73 graus Celsius, com as áreas mais quentes nas regiões litorânea, central e norte do estado, de acordo com dados das normais climatológicas de 1981 a 2010, ilustradas na Figura 22.

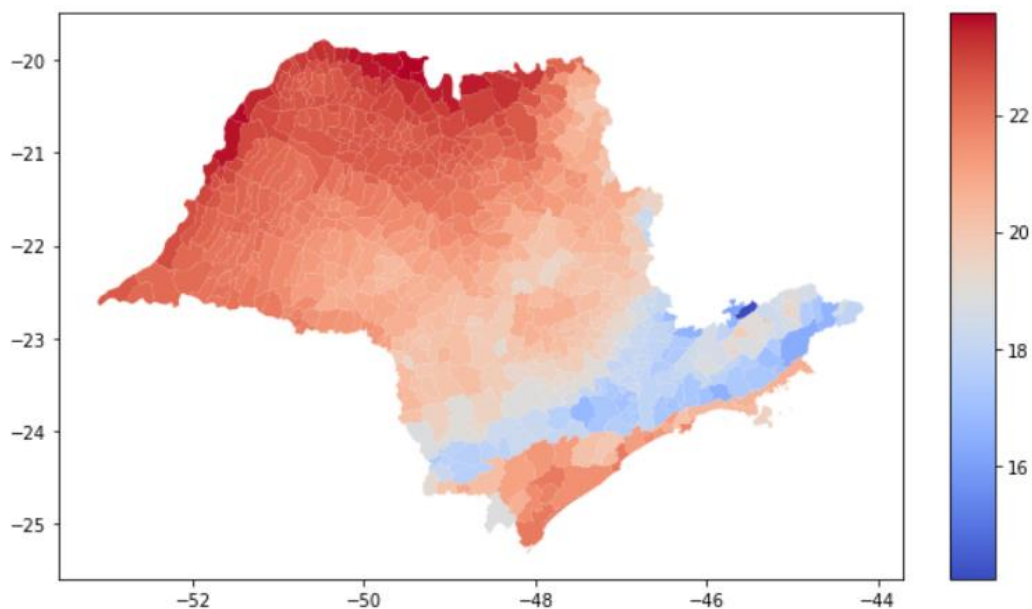


Figura 22 -Temperatura média anual para os municípios do estado de São Paulo.

Fonte: Dados: Normais Climatológicas (FICK; HIJMANS, 2017).

5.1.3 Infraestrutura de Saúde

Segundo dados disponibilizados pelo Departamento de Informática do Sistema Único de Saúde (DATASUS), o estado de São Paulo teve, em dezembro/2021, 94.158 leitos de internação, sendo 58,6% do Sistema Único de Saúde (DATASUS, 2022a). Considerando o total de leitos de internação, são 2,02 leitos para cada 1000 habitantes, abaixo da taxa nacional, que foi de cerca de 2,10 em dezembro de 2021.

Considerando as unidades de atendimento básicas (postos de saúde, unidades básicas e unidades de saúde da família), existia 0,12 unidades para cada mil habitantes no estado (dezembro/2021), quase metade da taxa nacional, que foi de 0,23 unidades para cada mil habitantes, de acordo com dados do Cadastro Nacional de Estabelecimentos de Saúde (DATASUS, 2022a).

Em relação aos profissionais de saúde, o estado tinha em dezembro/2021 cerca de 2,79 médicos e 1,48 enfermeiros para cada mil habitantes, acima dos valores nacionais para o mesmo período, de 2,17 médicos e 1,28 enfermeiros para cada mil habitantes. Para os agentes comunitários de saúde, a taxa para São Paulo era de 0,73 agentes para cada mil habitantes, abaixo da taxa nacional, de 1,33 agentes para cada mil habitantes (DATASUS, 2022b).

Observa-se que, enquanto os valores do estado para médicos e enfermeiros foram superiores aos valores nacionais, o mesmo não foi observado para dois componentes fundamentais na atenção básica: agentes comunitários de saúde e unidades básicas de saúde.

5.1.4 Investimentos

Segundo dados do Sistema de Informações sobre Orçamentos Públicos em Saúde (SIOPS), foram investidos em 2021 mais de 27 bilhões de reais em saúde no estado de São Paulo, cujos valores por subfunção são discriminados na Tabela 10 (MINISTÉRIO DA SAÚDE, 2022f). Com exceção dos investimentos na Subfunção Alimentação e Nutrição, os valores mais baixos investidos estão nas subfunções mais relacionadas ao combate à dengue: Atenção Básica, Vigilância Sanitária e Vigilância Epidemiológica.

Tabela 10 – Distribuição dos investimentos em saúde em SP em 2021.

Item	Valor	% Total
Atenção Básica	504.337.004,28	1,83
Assistência Hospitalar e Ambulatorial	23.037.426.818,01	83,78
Suporte Profilático e Terapêutico	1.860.099.252,41	6,76
Vigilância Sanitária	12.352.993,34	0,04
Vigilância Epidemiológica	359.119.696,59	1,31
Alimentação e Nutrição	132.411.684,60	0,48
Outras Subfunções	1.592.251.857,71	5,79
Total	27.497.999.306,94	100

Fonte: (MINISTÉRIO DA SAÚDE, 2022f)

Em relação ao investimento em cada subfunção por habitante, observa-se que, nas três mais relacionadas ao combate à dengue, os investimentos do estado de São Paulo são inferiores aos demais estados da Região Sudeste, conforme demonstrado na Tabela 11. Observa-se que, em relação às subfunções relacionadas à assistência, SP tem maior investimento per capita do que os outros estados, enquanto, para subfunções mais relacionadas à prevenção, o estado de SP tem menores investimentos por habitante.

Tabela 11 - Investimentos em saúde por habitante em SP e demais estados da Região Sudeste em 2021.

Subfunção	R\$/Hab – São Paulo *	R\$/Hab – Sudeste **
Atenção Básica	10,81	44,76
Assistência Hospitalar e Ambulatorial	493,84	413,33
Suporte Profilático e Terapêutico	39,87	22,92
Vigilância Sanitária	0,26	4,11
Vigilância Epidemiológica	7,7	14,2
Alimentação e Nutrição	2,84	0,01
Outras Subfunções	34,13	87,84
Total	589,46	587,17

* População estimada em 2021 (IBGE): 46.649.132

** População estimada em 2021 (IBGE): 42.983.780

Fontes dos dados: (MINISTÉRIO DA SAÚDE, 2022f) e População Estimada (IBGE, 2022d)

5.1.5 Dengue

Em São Paulo, estado mais populoso do Brasil, a infestação pelo mosquito *Ae. aegypti* estava concentrada da região noroeste do estado em 1986 e, 30 anos depois, já havia se espalhado para 99,69% dos municípios paulistas, não apenas pelo *Ae. aegypti*,

mas também por outro vetor, o *Aedes albopictus*, conforme ilustrado na Figura 23 (FONSECA JÚNIOR et al., 2019).

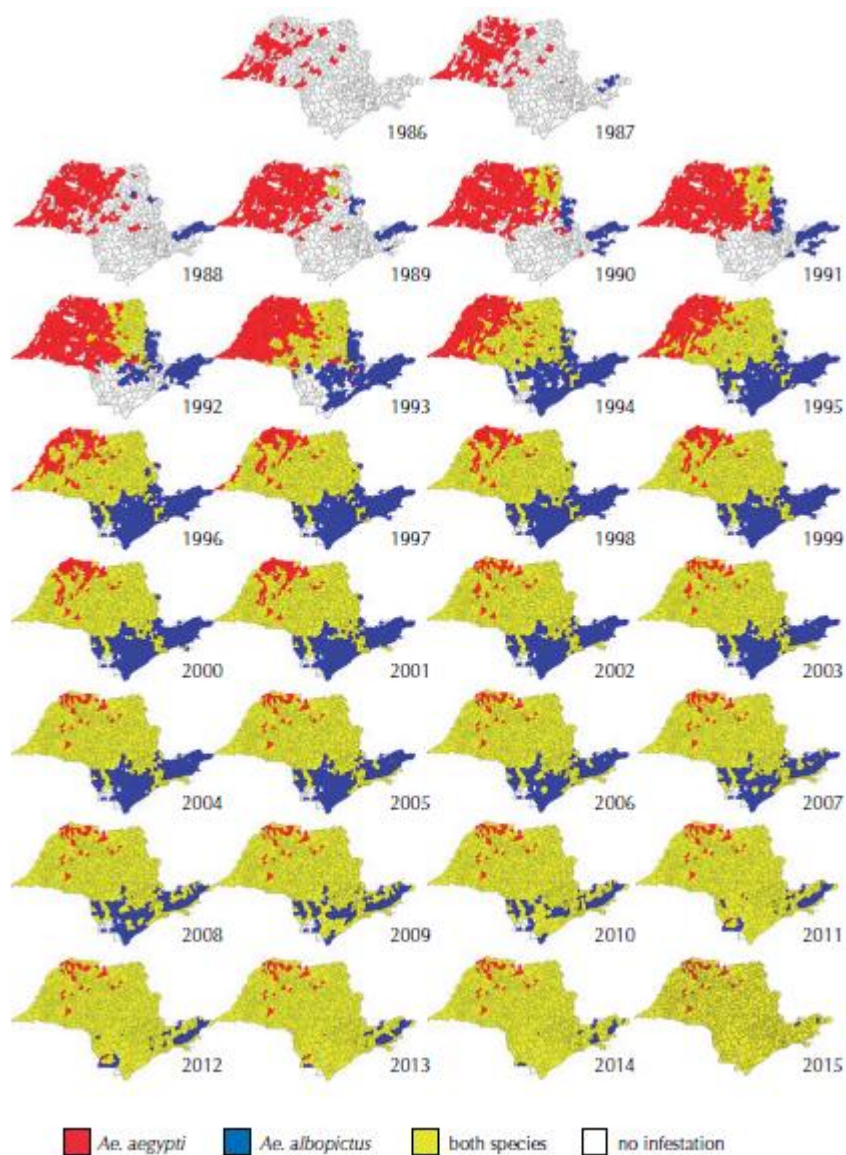


Figura 23 - Municípios do estado de São Paulo infestados pelas duas espécies transmissoras do vírus da dengue.

Fonte: (FONSECA JÚNIOR et al., 2019)

Segundo dados do SINAN, foram registrados 160.414 casos prováveis de dengue no estado de São Paulo em 2021, com uma incidência de 343,87 casos / 100 mil habitantes. Esse valor é superior ao encontrado para os demais estados do Brasil no mesmo período, que registrou 230,61 casos / 100 mil habitantes, e mais de cinco

vezes superior a taxa de incidência dos demais estados da Região Sudeste, que foi de 62,59 casos / 100 mil habitantes (SINAN, 2022).

Analisando-se a série histórica de casos de 2001-2021, em geral a taxa de incidência do estado de São Paulo esteve abaixo do valor encontrado para os demais estados da Região Sudeste (16 em 21 anos). No entanto, nos últimos dois anos da série, a incidência em São Paulo superou o valor encontrado para a região Sudeste e, em 2021, a incidência no estado foi superior à taxa de incidência para o Brasil, fato que já havia ocorrido em outros cinco anos, conforme Figura 24.

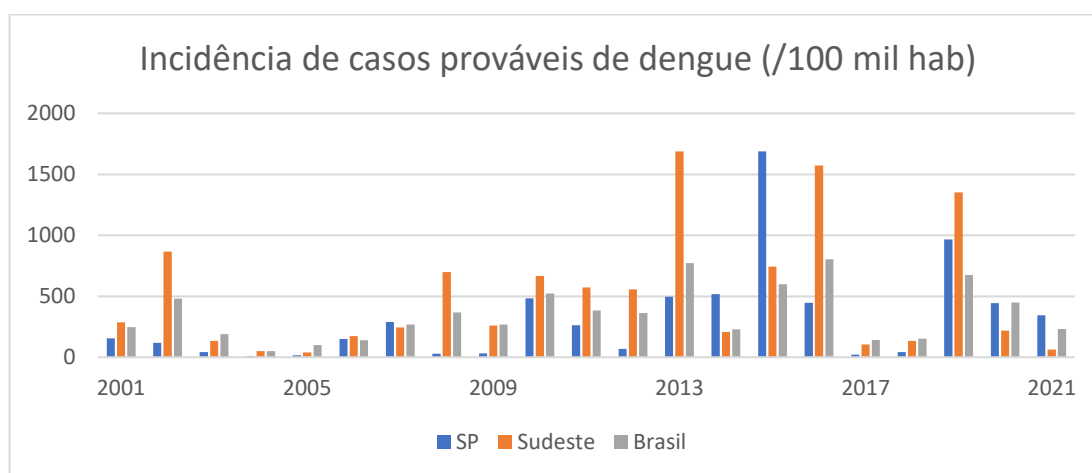


Figura 24 - Taxa de incidência de casos prováveis de dengue para o estado de SP, Sudeste (sem considerar SP) e Brasil (sem considerar SP).

Fonte: Dados: (SINAN, 2022)

Quando se analisa a distribuição espacial da taxa de incidência dentro do estado de São Paulo, é possível observar que as maiores taxas estão presentes nas regiões com temperaturas mais altas no estado, conforme Figura 25.

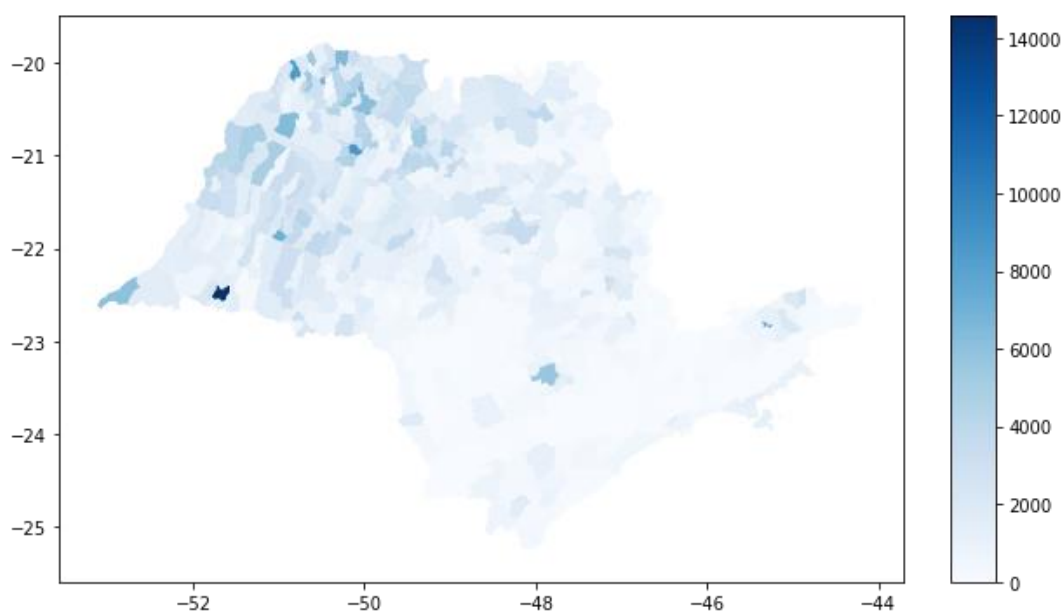


Figura 25 -Taxa de incidência para os municípios do estado de SP. Dados: média móvel de casos prováveis 2019-2021.

Fonte: Dados: (SINAN, 2022).

Como um baixo número de casos pode gerar uma alta taxa de incidência em municípios com populações muito pequenas, uma visualização alternativa pode considerar a incidência em função do porte populacional, conforme indicado nas diretrizes da Secretaria de Saúde do Estado de São Paulo (SECRETARIA DE SAÚDE DE SÃO PAULO, 2017). Nesse caso, a Figura 26 ilustra a incidência de casos prováveis (2019-2021) em função do coeficiente por porte populacional descrito na Tabela 5. Embora o padrão em relação a temperatura tenha se mantido, é possível observar mais claramente alguns municípios cuja incidência tenha tido altos valores em relação ao valor esperado pelo porte populacional.

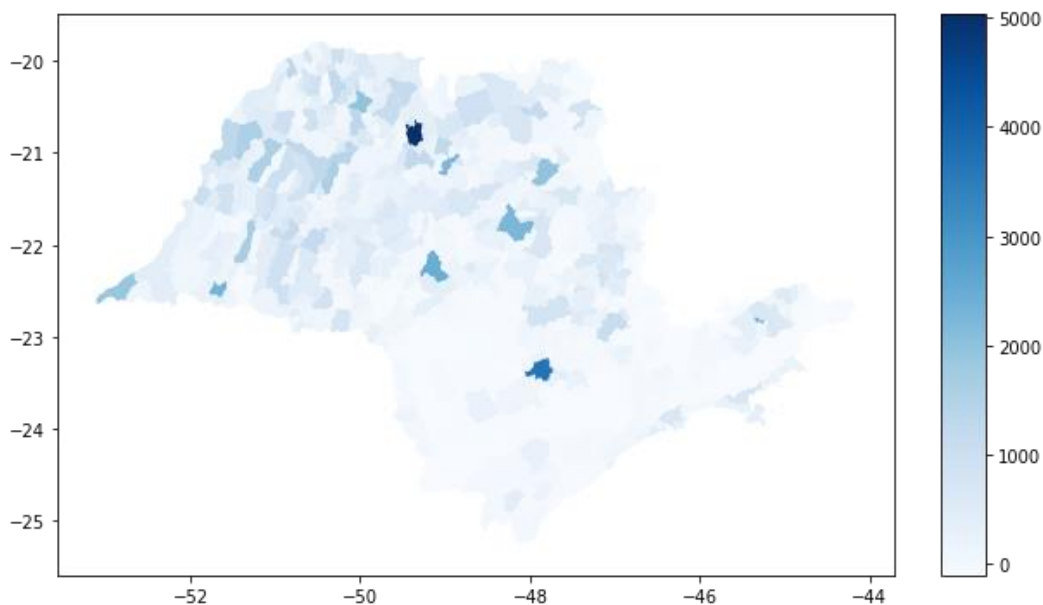


Figura 26 - Taxa de incidência para os municípios do estado de SP em função do coeficiente por porte populacional.

Fonte: Dados: média móvel de casos prováveis 2019-2021 (SINAN, 2022).

5.2 Descrição dos Experimentos

A definição dos experimentos para avaliação da metodologia proposta foi realizada em conjunto com os especialistas de domínio, considerando principalmente dois objetivos: 1) demonstrar o potencial da metodologia proposta para avaliação em saúde, e 2) atender demandas atuais para a gestão do programa de combate à dengue.

A seguir, os dois experimentos são descritos em detalhes, incluindo os passos para suas realizações e resultados esperados.

5.2.1 Experimento 1: Avaliação Donabedian de acordo com o contexto

5.2.1.1 Objetivo

Avaliar o programa de combate à dengue em todos os municípios do estado de São Paulo usando o modelo de Donabedian e considerando o contexto local.

5.2.1.2 Descrição

Neste experimento, o programa de combate à dengue será avaliado nos municípios considerando os componentes do modelo de avaliação em saúde de Donabedian, além do contexto em que o programa é aplicado.

Inicialmente, serão aplicados métodos supervisionados de aprendizado de máquina para seleção das variáveis de contexto mais relacionadas aos casos de dengue. Em seguida, os municípios serão agrupados de acordo com as variáveis selecionadas no passo anterior. Para esta tarefa, serão executados algoritmos não-supervisionados para identificar os grupos de municípios. Uma vez definidos os grupos, os municípios de cada grupo serão avaliados quanto aos componentes definidos no modelo de Donabedian: estrutura, processo e resultado.

5.2.1.3 Justificativa

A avaliação do programa em um grande conjunto de municípios pode ser custosa e demorada, devido ao custo para coleta de dados e dificuldades na integração dos dados de várias fontes. Além disso, devido às particularidades de cada município, uma avaliação comparativa de resultados deveria considerar o contexto em que o programa é aplicado, especialmente nos aspectos que podem ter influência na disseminação da dengue, como clima, uso do solo e economia.

5.2.1.4 Contribuições Esperadas

Em primeiro lugar, a agregação de métodos de Ciência de Dados aos processos de avaliação já estabelecidos pode aumentar o potencial de descoberta de conhecimento no cenário de *Big Data*. Além disso, a avaliação, conforme as particularidades locais, pode auxiliar os gestores de saúde na identificação eficiente dos locais com maior risco, considerando cenários atuais e futuros como, por exemplo, o avanço da urbanização e as mudanças climáticas. Por fim, possíveis diferenças de resultados entre municípios de mesmo grupo (contexto similar) podem ser úteis em novas pesquisas para identificar as razões pelas quais municípios similares alcançam resultados diferentes.

5.2.1.5 Atividades

Este experimento foi desenvolvido por meio das seguintes atividades:

- Definição do modelo de avaliação: foi construída a matriz de avaliação, com definição dos indicadores adotados por componente e identificação dos conjuntos de dados necessários;
- Aquisição dos dados: os conjuntos de dados foram coletados junto a repositório de dados;
- Seleção das variáveis: foram executados algoritmos supervisionados para identificação e seleção das principais variáveis de contexto relacionadas à dengue;
- Identificação dos grupos: através da execução de algoritmos de clusterização, foram identificados os grupos de municípios com características similares;
- Aplicação da avaliação: para cada município foi aplicado o modelo de avaliação definido no Passo 1, considerando a pontuação do município para cada indicador
- Geração de resultados: de acordo os resultados dos passos anteriores, as seguintes questões foram respondidas:
 - Quais são as principais variáveis de contexto relacionadas aos casos de dengue?
 - Quais são as principais diferenças quanto a contexto e casos entre os grupos?
 - Existem diferenças de resultados entre municípios de cada grupo? Se sim, quais diferenças foram encontradas nos indicadores de estrutura e processo avaliados?

5.2.2 Experimento 2: Painel BI para avaliações baseadas no modelo de Donabedian

5.2.2.1 Objetivo

Propor um método para avaliação contínua do programa segundo os componentes do modelo proposto por Avedis Donabedian.

5.2.2.2 Descrição

Para este experimento, foi proposto um painel para avaliação dos indicadores dos municípios do estado de São Paulo quanto aos componentes do modelo de Donabedian. A proposta desse painel é que seja uma ferramenta para avaliação contínua dos municípios do estado, possibilitando aos gestores a visualização rápida da situação do programa no estado.

Para isso, foi desenvolvido um processo de integração contínua dos dados necessários para geração dos indicadores, que foram armazenados em um *Data Warehouse*. A partir dessa estrutura, foi desenvolvido um painel interativo para visualização dos indicadores, onde os gestores poderão avaliar os resultados de acordo com aspectos espaciais e temporais.

5.2.2.3 Justificativa

A avaliação em saúde é, normalmente, um processo com alto custo para a sua realização. No entanto, a execução contínua pode ser fundamental para identificação de possíveis problemas e consequente tomada de decisão no momento adequado, ou seja, visando ações de prevenção. Com o grande número de municípios do estado (645) e as dificuldades para avaliação de cada um, a visualização integrada dos indicadores pode ser fundamental para a gestão estadual realizar a verificação contínua do que está sendo realizado pelos municípios e, assim, realizar as intervenções de forma mais rápida.

5.2.2.4 Contribuições Esperadas

O painel proposto neste experimento pode ser uma importante ferramenta para suporte à gestão do programa de combate à dengue, provendo informações rápidas aos usuários. Além disso, espera-se que a arquitetura proposta para este painel possa ser aplicada na avaliação em saúde de outras regiões ou em outros programas de saúde.

5.2.2.5 Atividades

Este experimento foi organizado nas seguintes etapas:

- Definição do painel: foi definido, junto aos especialistas de domínio, quais indicadores poderão ser visualizados no painel;
- Criação do *Data Warehouse*: foi criada a estrutura de dados que será utilizada pelo painel utilizando-se um banco de dados relacional;
- Aquisição dos dados: os conjuntos de dados foram coletados junto à repositório de dados e, após as transformações necessárias, inseridos no *Data Warehouse*;
- Criação do painel: para demonstração do experimento, foi construído o painel usando uma ferramenta de *Business Intelligence* (BI) consolidada.

6 RESULTADOS

A partir do objetivo geral desta pesquisa, a avaliação do programa de combate à dengue, são descritos a seguir os resultados dos demais componentes da metodologia proposta no Capítulo 4.

De acordo com a metodologia proposta, um repositório comum de conjuntos de dados é necessário para a execução das avaliações descritas no capítulo anterior do programa de controle da dengue. Para atender a esses requisitos, o repositório foi criado de acordo com o esquema ilustrado na Figura 27. Nesse esquema, detalhado nas subseções a seguir, são identificados os artefatos computacionais adotados para as três partes necessárias para realização das avaliações, sendo 1) fontes de dados; 2) repositório de dados e 3) avaliações do programa com objetivos específicos.

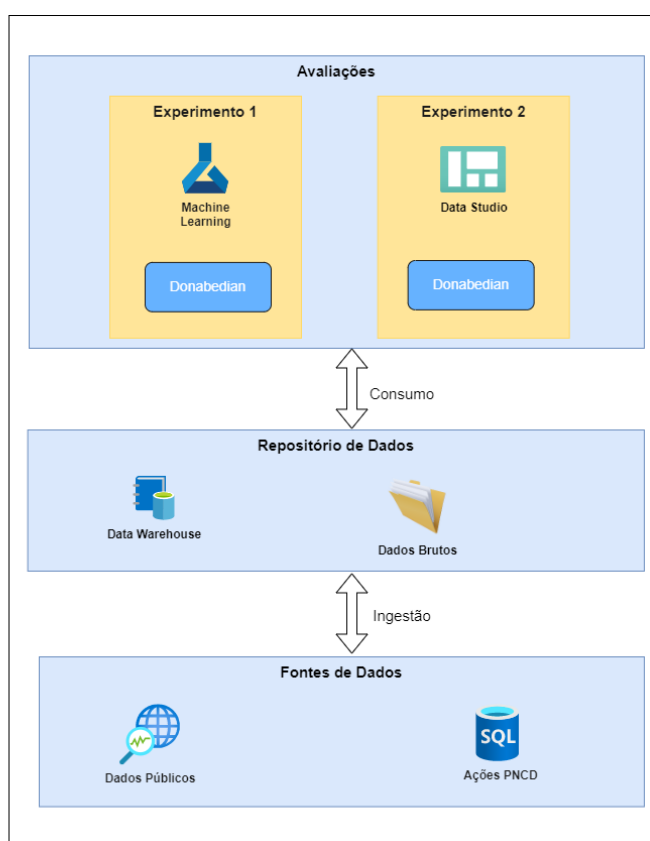


Figura 27 – Arquitetura definida para execução das avaliações.

6.1 Fontes de Dados

Para esta pesquisa, foram utilizados dados de duas origens:

- Fontes de dados públicas relevantes ao problema da dengue: conjuntos de dados disponíveis de diversos sistemas de acesso público, como casos de dengue do SINAN, dados sociodemográficos, clima, entre outros;
- Ações de combate à dengue: banco de dados gerenciado por um órgão da Secretaria de Saúde do Governo do Estado de São Paulo, que contém dados das ações de combate à dengue executadas pelos municípios.

Tabela 12 - Fontes de dados necessárias para o desenvolvimento da pesquisa.

Categoria Variável	Variáveis	Fonte de Dados
Contexto - Clima	Temperatura Média Anual Precipitação	(FICK; HIJMANS, 2017)
Contexto - Condições socioeconômicas	IDH	(IPEA, 2022)
	PIB	(IBGE, 2022e)
Contexto - Urbanização	População Urbana	(SISTEMA NACIONAL DE INFORMAÇÕES SOBRE SANEAMENTO (SNIS), 2022)
	Uso do solo	(SOUZA et al., 2020)
Contexto - Geral	Área total	(SOUZA et al., 2020)
	População	(IBGE, 2022d)
Estrutura	Número de ACS e ACV Total de médicos Total de enfermeiros Total de equipes de saúde da família Unidades básicas de saúde	(DATASUS, 2022a, 2022b)
	Despesas em saúde	(MINISTÉRIO DA SAÚDE, 2022f)
	Despesas em saúde com recursos próprios	(MINISTÉRIO DA SAÚDE, 2022f)
Processo	Visitas casa a casa	(SECRETARIA DE ESTADO

	Controle de criadouros ADL Visitas Imóveis Especiais e Pontos Estratégicos	DA SAÚDE DE SÃO PAULO, 2022)
Resultado	Casos de dengue	(SINAN, 2022)
	Índice de Infestação	(SECRETARIA DE ESTADO DA SAÚDE DE SÃO PAULO, 2022)

Para cada fonte de dados selecionada, foram identificados os mecanismos de acesso para obtenção dos dados. Todos os dados listados na Tabela 12 estão em escala de município e se referem ao período de 2015 a 2019, com exceção dos dados climáticos, disponíveis até 2018, e uso do solo, disponíveis de 1985 a 2021. Os dados dos tipos Contexto, com exceção de clima e uso do solo, e Estrutura foram coletados por meio de filtros disponíveis nos portais de acesso e os resultados armazenados em arquivos no formato separado por vírgulas (.csv). Os dados de clima foram disponibilizados em formato *raster* (.tiff), enquanto que os dados de uso do solo foram obtidos em formato de planilha (.xlsx).

Para as variáveis da categoria Processo, os dados foram selecionados a partir do banco de dados e armazenados em formato *csv*. Por fim, para a categoria Resultado, os dados de casos de dengue foram obtidos do portal do DATASUS e armazenados em formato *csv*, enquanto que para a variável de índice de infestação, os dados foram obtidos a partir do banco de dados da secretaria de saúde e armazenados em formato *csv*.

6.2 Repositório de Dados

A partir das informações coletadas na segunda fase e do modelo de avaliação definido na terceira fase, foi definida a arquitetura para o repositório de dados a ser utilizado pelas atividades de avaliação do programa. Essa arquitetura foi baseada nos conceitos de *Lakehouse*, que integra os benefícios de *Data Warehouses* com os das soluções de *Data Lake*. Esse modelo arquitetural, que pode ser utilizado para avaliação de outros programas ou intervenções da área da saúde, não está vinculado a um

fornecedor ou tecnologia específica, podendo ser implementado de acordo com as características da instituição.

Para desenvolvimento do repositório de dados, foram utilizados dois recursos de computação em nuvem de acordo com o estágio dos conjuntos de dados (brutos e transformados):

- Armazenamento de dados brutos: os dados extraídos das diversas fontes descritas anteriormente foram armazenados no serviço de armazenamento do Google Cloud. Nesse ambiente de armazenamento, os arquivos foram armazenados em seus formatos originais;
- Armazenamento de Dados Transformados: para utilização dos dados pelos dois experimentos, uma estrutura de banco de dados foi criada. Nessa estrutura, foram armazenados os dados após diversas atividades de transformação realizadas durante cada experimento. A plataforma utilizada para armazenamento dos dados transformados foi o Google BigQuery.

As atividades de ingestão para armazenamento dos dados brutos foram realizadas através dos mecanismos disponíveis pela plataforma em nuvem. Para o armazenamento dos dados transformados, os detalhes relativos à seleção, transformação e ingestão são descritos a seguir para cada experimento.

6.3 Avaliações

Como os dois experimentos utilizaram-se dos mesmos indicadores para avaliação, foi desenvolvida uma matriz única de avaliação. Em relação aos conjuntos de dados necessários, como foram analisados aspectos diferentes em cada experimento, as atividades de seleção e transformação dos dados foram específicas para cada experimento.

No caso do Experimento 1, avaliação do tipo pontual, os dados foram obtidos apenas uma vez. Dessa forma, a atividade de busca e seleção dos conjuntos de dados necessários foi realizada pontualmente, sem necessidade de execuções recorrentes. Os dados foram transformados e armazenados em uma estrutura temporária de banco de dados para execução das atividades de análise.

Em avaliações realizadas de forma contínua, como a que foi desenvolvida no Experimento 2, os dados devem ser obtidos constantemente junto ao repositório, o que pode ser feito de forma periódica ou sempre que um conjunto de dados novo está disponível. No caso desta pesquisa, o acesso foi realizado de forma periódica no Experimento 2, para fins de validação da solução proposta.

6.3.1 Matriz de Avaliação

A matriz de avaliação (Tabela 13) foi desenvolvida de acordo com os componentes de Donabedian e contendo 16 indicadores, sendo: oito indicadores para o componente Estrutura, quatro indicadores para o componente Processo e quatro para o componente Resultado. A seleção desses indicadores foi realizada de acordo com a literatura existente e com o apoio de um comitê de especialistas.

Para comparações entre os municípios, os indicadores foram calculados em função de uma característica comum. Nos indicadores referentes a Estrutura, os valores foram calculados em função da população. No caso dos dois indicadores relacionados à investimentos em saúde, foi calculada a razão simples do valor do investimento no ano em relação ao total de habitantes no período. Para os demais indicadores de Estrutura, os valores foram calculados em para cada 1000 habitantes.

Para os indicadores de Processo, foram utilizadas normalizações distintas. As quantidades de ações de visitas casa a casa e o controle de criadouros foram normalizados em função do número de quarteirões da cidade. Em relação às visitas a imóveis especiais e pontos estratégicos, os cálculos dos indicadores foram realizados de acordo com o número total de imóveis desses tipos. Para a avaliação de densidade larvária foram utilizados os valores totais.

Em relação aos indicadores de Resultados foram adotados, com exceção do número de semanas com casos, as formas normalmente utilizadas para cálculo desses valores, que já permitem comparação entre locais diferentes. Os dois indicadores de incidência são calculados em função do número de casos por 100000 habitantes, enquanto o índice de infestação é uma razão entre o número de imóveis positivos para o número de imóveis trabalhados. Essa matriz de avaliação foi, ao final, validada junto

ao comitê de especialistas. Essa validação é importante para garantir a adequação do modelo avaliativo às características do objeto avaliado.

Tabela 13 - Matriz de avaliação para dengue segundo componentes do modelo de Donabedian.

Indicador	Descrição	Fórmula
Estrutura		
Quantidade de UBS	Número de unidades básicas de saúde do município	(Total de UBS / população) * 1000
Quantidade de Médicos	Número de médicos no município	(Total de Médicos / população) * 1000
Quantidade de Enfermeiros	Número de enfermeiros no município	(Total de Enfermeiros / população) * 1000
Quantidade de Equipes de Saúde da Família	Número de equipes de saúde da família no município	(Número de equipes / população) * 1000
Investimentos em Saúde – Recursos próprios	Valor total de recursos próprios investidos em saúde	Valor investido / população
Despesas em saúde	Total de despesas em saúde	Valor investido / população
Quantidade de ACS	Número de agentes comunitários de saúde	(Total de ACS / população) * 1000
Quantidade de ACE	Número de agentes de combate à endemia	(Total de ACE / total de imóveis) * 1000
Processo		
Visita imóveis	Quantidade de visitas a imóveis no período	Total de visitas / número de bairros
Controle de criadouros	Quantidade de ações de controle de criadouros executadas no período.	Total de ações / número de bairros
ADL	Total de ADLs realizadas	-
PEs e IE	Visitas em imóveis especiais/pontos estratégicos	Razão pelo total de imóveis do município
Resultado		
Incidência de casos de dengue por porte do município	Razão da incidência pelo coeficiente de incidência por porte de município	Incidência anual / coeficiente por porte de município
Incidência de casos de dengue	Número de casos de dengue no período	(Total de casos de dengue / população) * 100000
Semanas epidemiológicas com casos notificados	Número de semanas epidemiológicas com casos notificados de dengue	-
Índice de infestação	Índice Predial de infestação	(Imóveis positivos / imóveis pesquisados) * 100

6.4 Experimento 1: Avaliação Donabedian de acordo com o contexto

6.4.1 Descrição do experimento

Para avaliação dos resultados desse experimento, foram definidas as seguintes questões de pesquisa:

- Quais são as principais variáveis de contexto relacionadas aos casos de dengue?
- Quais são as principais diferenças entre os grupos quanto a contexto e número de casos?
- Existem diferenças de resultados entre municípios de cada grupo? Se sim, quais diferenças foram encontradas nos indicadores de estrutura e processo avaliados?

6.4.2 Etapas

O experimento foi desenvolvido em quatro etapas, desde a obtenção dos dados até a avaliação de acordo com o modelo de Donabedian. A primeira etapa consistiu das atividades relacionadas a aquisição dos dados necessários junto ao repositório de dados. Nessa etapa, os dados foram selecionados de acordo com informações coletadas junto ao painel de especialistas e à literatura existente sobre a dengue. Assim, dois aspectos puderam ser definidos:

- Período de seleção dos dados;
- Conjunto de dados necessários.

Em relação ao período, foi definido o intervalo entre os anos de 2015 e 2019 contemplando, quando possível, dados mensais e por semana epidemiológica. A escolha de um intervalo de cinco anos explica-se para diminuir a influência de um período com grande variação nos indicadores, como anos epidêmicos. Além disso, dada a sazonalidade da dengue, um intervalo maior pode incluir tanto anos com alto

número de casos assim como os anos subsequentes, em que a taxa de incidência tende a ser menor.

Em relação aos anos escolhidos, optou-se por anos mais recentes possíveis, devido a maior aderência às condições climáticas atuais e também pelo fato de que dados mais recentes podem incluir informações mais completas. Um exemplo disso é a informação da semana epidemiológica, dado fundamental para um dos indicadores. De acordo com dados públicos do SINAN disponíveis no portal do DATASUS, a informação da semana epidemiológica só está disponível a partir de 2014. Os anos de 2020 e 2021, embora tivessem dados disponíveis para boa parte dos indicadores, não foram selecionados pela ausência de um grupo de dados considerado altamente relacionado aos casos de dengue: clima. Outro ponto que influenciou na exclusão de 2020 e 2021 do filtro foi em relação a possíveis impactos que a pandemia da COVID-19 possa ter causado na execução das atividades de combate à dengue, visto que nesses anos boa parte das equipes de saúde tinha como prioridade ações relacionadas à pandemia. Nada impede, no entanto, que esses dados possam ser utilizados em outros experimentos que busquem avaliar o impacto da COVID-19 nas ações de combate à dengue.

O outro aspecto definido junto aos especialistas de domínio foi em relação aos conjuntos de dados necessários de acordo com a matriz de indicadores definida anteriormente. Para isso, selecionou-se conjunto de dados de acordo com sua categoria (contexto, estrutura, processo e resultado) e disponibilidade de acordo com o período definido. Uma vez que os dados foram selecionados, atividades de pré-processamento foram realizadas para preparação dos dados para as etapas seguintes do experimento.

A segunda etapa consistiu na execução de algoritmos de clusterização para identificação dos possíveis grupos de acordo com as variáveis de contexto. Para essa etapa foram avaliados diversos algoritmos, assim como foi analisado o melhor conjunto de atributos a ser utilizado pelos algoritmos. Essa análise foi realizada de acordo com critérios definidos para comparação dos resultados entre algoritmos e conjuntos de atributos definidos.

Uma vez que os grupos foram identificados, foram executados, na terceira etapa, algoritmos de aprendizado de máquina para identificação das principais

variáveis utilizadas para classificar os grupos de acordo com o contexto. O objetivo desse tipo de análise é simplificar o processo de identificação dos grupos, pois apenas as variáveis mais importantes poderiam ser selecionadas em futuras análises.

Por fim, na quarta etapa, o modelo de Donabedian foi aplicado para análise dos componentes (estrutura, processo e resultado) nos grupos identificados nos passos anteriores. Essa análise foi realizada tanto na comparação dos grupos entre si assim como a análise interna de cada grupo.

6.4.3 Resultados

6.4.3.1 Etapa 1: Seleção e preparação dos dados

Os conjuntos de dados foram selecionados no repositório de dados com o intervalo definido (2015-2019) e considerando as variáveis de acordo com a categoria (contexto, estrutura, processo e resultado). Além disso, algumas atividades de pré-processamento foram executadas para a integração dos dados, como extração e transformação de informações temporais e espaciais. Ademais, os dados transformados foram armazenados em um banco de dados para possibilitar a repetição das atividades sempre que necessário. Abaixo, são detalhadas as atividades conduzidas nessa etapa:

- Seleção dos conjuntos de dados: busca dos conjuntos de dados no repositório para o período definido. Os conjuntos de dados selecionados foram armazenados em uma área temporária para as atividades de pré-processamento;
- Análise exploratória: avaliação dos conjuntos de dados para identificar aspectos sobre o conteúdo, como dados ausentes, forma como os dados foram preenchidos. Essa atividade foi necessária para definir a necessidade de implementação de *scripts* de pré-processamento. Foi possível identificar, por exemplo, que as informações-chave para possível junção dos conjuntos não estavam preenchidas de maneira uniforme. Enquanto alguns conjuntos de dados tinham apenas o nome do município, outros conjuntos de dados apresentavam, além do nome do município, o código no município no IBGE e

o código do estado. Mesmo em relação ao código do município foram observadas diferenças, já que alguns conjuntos de dados continham o código em sete caracteres e outros em seis caracteres;

- Pré-processamento: desenvolvimento e execução de *scripts* para extrair os dados de interesse do experimento, descritos abaixo:
 - Extração das variáveis climáticas: os dados históricos de temperaturas máximas e mínima e precipitação estavam presentes em arquivos do tipo *raster* para todo o planeta. Dessa forma, para extração do valor de cada variável, foram executados scripts de geoprocessamento para 1) extrair os shapes correspondentes aos municípios do Estado de São Paulo e, 2) cálculo da mediana do município para a variável de interesse. Como os arquivos são disponibilizados pela fonte no formato ano-mês, os dois scripts foram executados para cada um dos arquivos (108 arquivos para cada variável, no período entre 2010-2018);
 - Extração das informações de uso do solo: os dados disponibilizados pelo MapBiomas (SOUZA et al., 2020) contém medidas de área para cada classe, como áreas urbanas, agricultura, rios, etc. Por isso, um script foi necessário para extração apenas da área correspondente a ações antropogênicas para cada município/ano;
 - Extração das demais variáveis de interesse: para todos os conjuntos de dados foram extraídos os valores correspondentes a 1) identificador do município; 2) variáveis temporais (ano, mês, semana epidemiológica) e 3) variável quantitativa, que correspondia ao que estava sendo disponibilizado no conjunto de dados (ex.: número de casos de dengue, população, PIB, etc.). Para essa última, os valores faltantes foram preenchidos de acordo com as características dos campos e fonte. Para dados oriundos do portal do DATASUS, os valores nulos são disponibilizados com o caractere “-“. Nesse caso, os valores foram preenchidos com o valor zero. Em outros casos, os valores faltantes foram preenchidos através de interpolação. Os dados pré-processados

foram, por sua vez, armazenados na área temporária para outras análises.

- Ingestão: Armazenamento dos dados transformados em um banco de dados no Google BigQuery. A adoção de um serviço na nuvem foi feita devido a facilidade de integração com as ferramentas de processamento em nuvem, utilizadas nesse experimento, e também pela integração com a ferramenta utilizada para execução da avaliação de acordo com o modelo de Donabedian. Além disso, o banco de dados foi estruturado de forma a separar os dados de acordo com a sua categoria (contexto, estrutura, processo e resultado).

Tabela 14 - Variáveis de contexto selecionadas para a etapa de clusterização.

Variável	Descrição
IDHM	Valor do IDH municipal para o ano de 2010
Temperatura máxima	Média da temperatura máxima anual
Temperatura mínima	Média da temperatura mínima anual
Temperatura média	Temperatura média anual
Precipitação	Total de chuvas (em mm) no ano
PIB per capita	Produto interno produto per capita
Mortalidade infantil	Taxa de mortalidade infantil (por mil nascimentos)
População urbana	Taxa de população urbana
Uso de solo	Porcentagem da área do município utilizada em atividades antropogênicas
Densidade	Densidade populacional
Abastecimento de água	Porcentagem de domicílios com abastecimento de água
Esgoto	Porcentagem de domicílios com rede de esgoto

6.4.3.2 Etapa 2: Clusterização

Nessa etapa, foram selecionados os dados da categoria Contexto a partir do banco de dados construído na etapa anterior. Além da variável identificadora do

município, foram selecionadas doze variáveis, conforme descrito na Tabela 14. Com exceção do IDH, foi calculado o valor médio de cada variável no período de 2015 a 2019.

Em seguida, foi utilizado o algoritmo *KMeans* para identificação dos grupos, que é amplamente utilizado em experimentos de clusterização. Uma característica desse algoritmo é a necessidade da definição a priori do número de clusters. Existem diversos métodos para identificação do número ótimo de clusters e um deles é o Método Elbow, em que o número ótimo de clusters é obtido através de análise gráfica (SYAKUR et al., 2018). O número ótimo de clusters é aquele onde a partir do qual não se observa grande mudança nos resultados. De acordo com a Figura 28, o número ótimo poderia ser quatro ou cinco. Como o número de municípios é relativamente pequeno (645), optou-se pelo menor número de clusters (4).

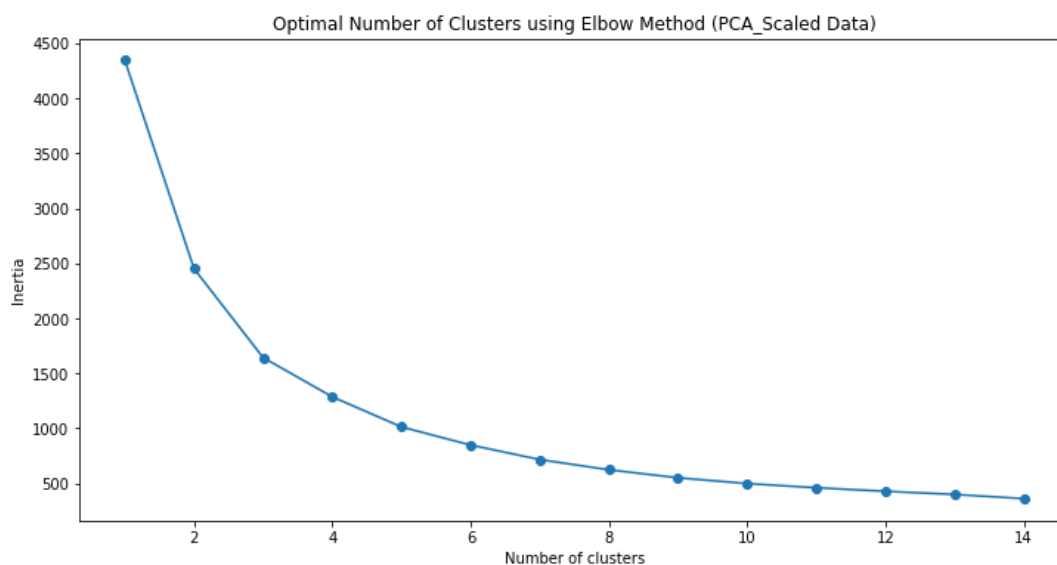


Figura 28 - Avaliação do número ótimo de clusters.

Ressalta-se que a mesma avaliação foi realizada tanto com a extração das componentes principais (Principal Component Analysis – PCA) como com dados brutos. Em ambos os casos foram encontrados os mesmos resultados quanto ao número ótimo de clusters. Em seguida, foram realizadas diversas execuções com diferentes combinações das variáveis selecionadas. Para seleção do conjunto ótimo de variáveis a serem incluídas na análise, foram adotados três critérios:

- Silhouette Score: métrica utilizada para medir a qualidade do método de clusterização, medindo, para cada elemento, aspectos como coesão (no próprio cluster) e separação (em relação aos demais clusters). Com valor entre -1 e 1, quanto maior o valor melhor é o método de clusterização utilizado;
- Melhor distribuição: para melhor comparação entre os clusters, foi definido que não houvesse uma grande concentração dos elementos em um cluster apenas. Assim, mesmo que fosse obtido um Silhouette Score mais alto, privilegiou-se os modelos com uma distribuição menos concentrada de elementos nos clusters;
- Visualização espacial: uma vez que os grupos foram identificados, os mesmos foram avaliados através de visualização espacial para verificar se havia coerência nos resultados encontrados pelo algoritmo. Essa avaliação foi feita em conjunto com especialistas de domínio.

Para seleção dos conjuntos de dados, foram feitas análises a respeito daquelas variáveis com maior correlação com a incidência de casos de dengue, excluindo-se aquelas a partir de um ponto de corte. Além disso, foi avaliada a exclusão das variáveis cuja correlação entre si fosse maior do que um determinado ponto de corte. No entanto, foi observado que a melhor combinação possível de acordo com os três critérios estabelecidos foi aquela em que não houve exclusão de nenhuma variável. Dessa forma, para identificação dos grupos foi utilizado a extração das componentes principais para as doze variáveis de entrada. A distribuição dos grupos é descrita na Tabela 15.

Tabela 15 - Distribuições dos registros entre os grupos.

Grupo	Quantidade de registros
0	366
1	113
2	100
3	65

Embora houvesse uma maior concentração dos registros no Grupo 0 (366 municípios), esse foi o menor número identificado nas diversas execuções do algoritmo. Em relação ao Silhouette Score, o valor encontrado foi de 0,41, considerado satisfatório para o experimento de acordo com os demais critérios estabelecidos.

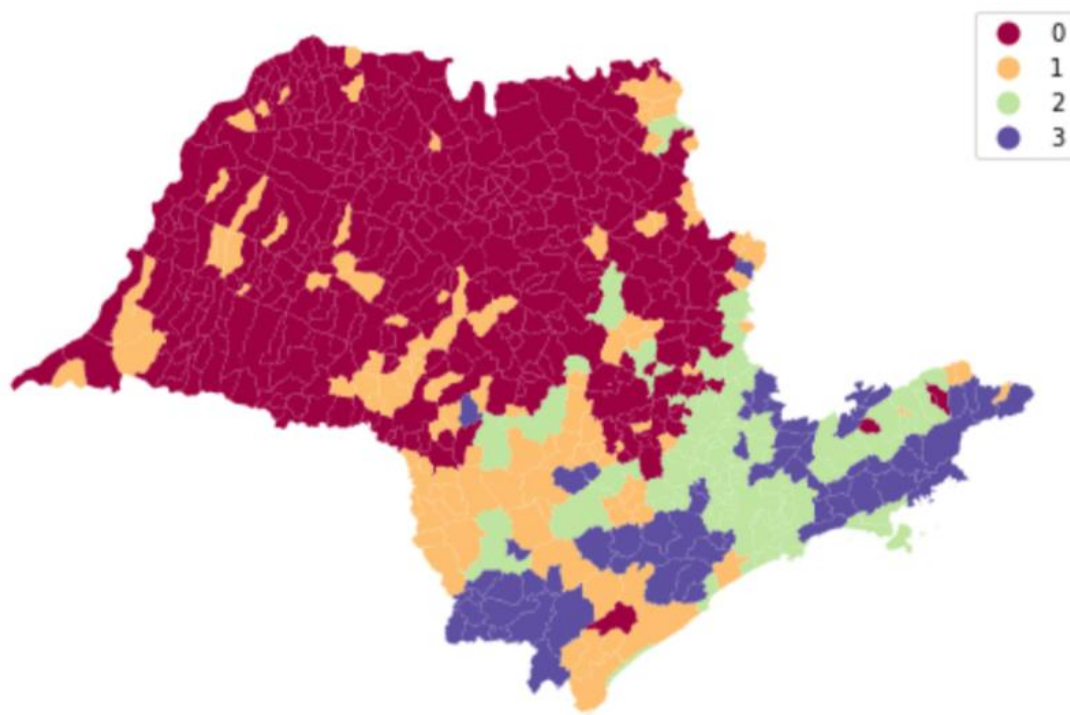


Figura 29 - Visualização espacial dos grupos identificados para os municípios do estado de São Paulo.

A visualização espacial dos grupos permitiu a verificação da coerência do modelo encontrado pelo algoritmo de acordo com padrões regionais. De acordo com a Figura 29, observa-se que os grupos 0 e 1 estão mais relacionados ao interior do estado, enquanto o grupo 2 concentra os municípios da região metropolitana de São Paulo. Já os municípios do grupo 3 estão concentrados no litoral e na região Sul do estado.

6.4.3.3 Etapa 3: Descrição das variáveis mais importantes

Após a identificação dos grupos, foi realizada a identificação das variáveis mais importantes na definição dos grupos. Para isso, foi utilizado o algoritmo Random

Forest, considerando o grupo como variável de classe e as doze descritas anteriormente como variáveis independentes. De acordo com a Figura 30, pode-se observar que a temperatura máxima foi a principal variável para decisão em relação aos grupos, o que é condizente com a visualização espacial ilustrada na Figura 29. As duas variáveis seguintes foram aquelas relacionadas a distribuição de água e saneamento. Essas três principais variáveis estão altamente relacionadas às condições para disseminação dos mosquitos transmissores da dengue.

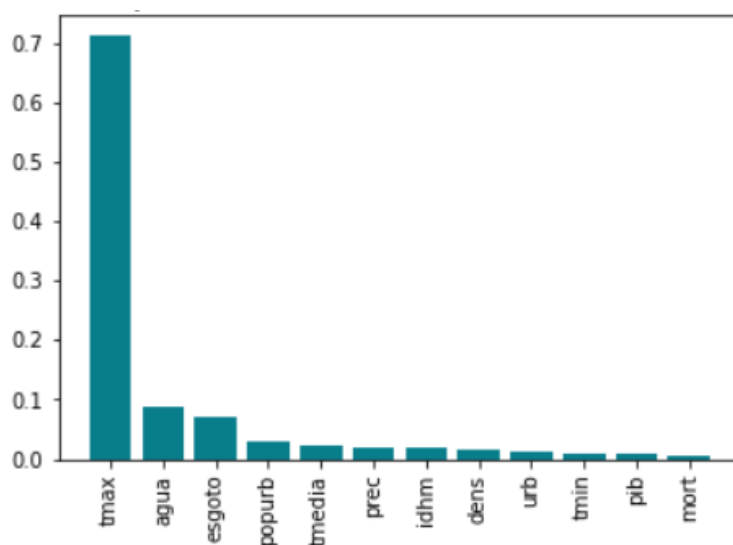


Figura 30 - Variáveis mais importantes para definição dos grupos de acordo com o algoritmo *Random Forest*.

No entanto, quando é feita a avaliação das variáveis mais importantes considerando a incidência como variável dependente, observa-se mudanças na ordem das variáveis, com exceção da temperatura máxima (Figura 31). A distribuição de água, segunda mais importante para definição dos grupos, foi a menos importante na relação a incidência de dengue. Já o PIB, segunda menos importante para definição dos grupos foi a terceira mais relacionada a incidência de dengue.

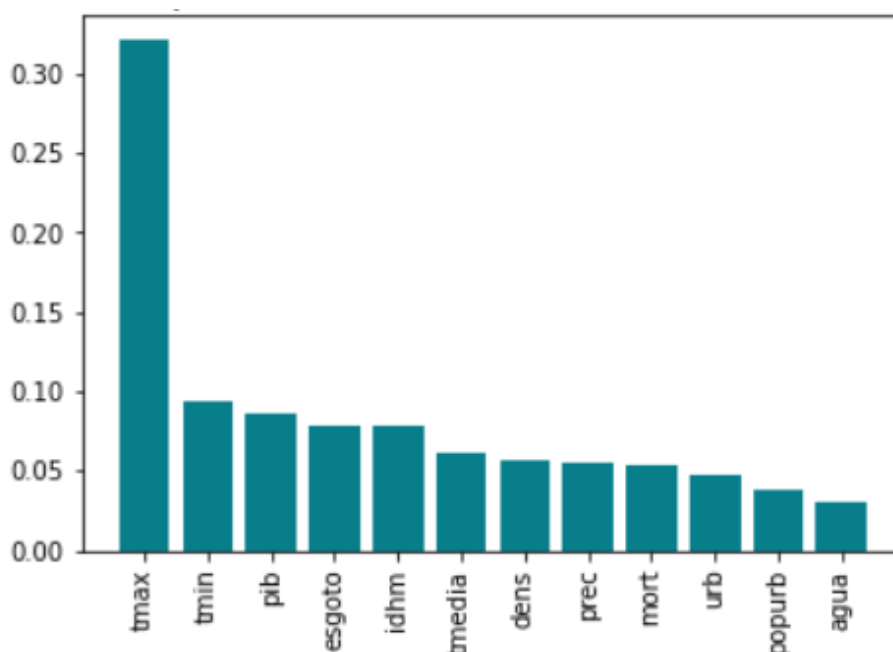


Figura 31 - Variáveis mais importantes considerando a incidência como variável dependente de acordo com o algoritmo *Random Forest*.

A identificação das variáveis mais importantes pode ser útil para simplificar o processo de clusterização, visto que nem sempre todas as variáveis estão disponíveis nos períodos desejados. Além disso, esse conjunto de variáveis pode ser utilizado em outros estudos para medir o peso que cada uma delas tem nos resultados de um indicador de saúde, como casos de dengue.

É importante observar que esse padrão pode ser diferente em outras unidades da federação, pois algumas características podem ter maior variabilidade do que no Estado de São Paulo.

6.4.3.4 Etapa 4: Análise dos resultados usando o modelo de Donabedian

Por fim, foram realizadas análises dos indicadores da matriz de avaliação em relação aos grupos. Para isso, foi utilizada a ferramenta Google Data Studio para acessar os dados armazenados no banco de dados criado para este experimento. Nas análises a seguir, inicialmente os grupos serão comparados entre si e, em seguida, será realizada a análise dentro de cada grupo.

6.4.3.4.1 Análise intergrupos

Na Tabela 16, são apresentados os valores médios de cada grupo de acordo com o componente Estrutura. Observam-se importantes diferenças entre os grupos, especialmente no grupo 2 para os demais. O número de agentes comunitários de saúde por 1000 habitantes (ACS), profissional fundamental para diversas ações de saúde, é bem abaixo dos demais grupos. O mesmo padrão é observado para outros indicadores relacionados a atenção primária em saúde, como número de agentes de controle de endemias (ACE), equipes de saúde da família (PSF) e número de unidades básicas de saúde (UBS). No entanto, o inverso ocorre em outras duas variáveis relacionadas a profissionais de saúde: número de médicos (Med) e enfermeiros (Enf) para cada mil habitantes. Uma possível explicação para essa diferença é que, de acordo com a visualização espacial dos grupos na Figura 29, os municípios desse grupo estão concentrados na área mais desenvolvida do estado que, por sua vez, concentra a maior parte das unidades de referência para saúde.

Em relação ao investimento em saúde, o Grupo 2 teve o segundo pior desempenho nas duas variáveis relacionadas: despesas totais em saúde (Despesa/hab) e despesas em saúde com recursos próprios (Desp Rec Próprios/hab).

Tabela 16 - Valores médios para os indicadores do componente Estrutura.

Grupo	ACE	ACS	Enf	Med	PSF	UBS	Despesa/hab	Desp Rec Próprios/hab
0	0,39	1,56	1,02	1,12	0,08	0,30	946,80	702,80
1	0,25	1,69	0,94	0,89	0,10	0,35	959,55	727,12
2	0,11	0,69	1,09	1,91	0,06	0,13	904,16	654,56
3	0,10	1,60	0,91	1,04	0,12	0,24	793,39	567,77

Quando são analisados os indicadores do componente Processo, listados na Tabela 17, nota-se que o Grupo 2 tem índice maior para dois dos quatro indicadores, com exceção de controle de criadouros e visitas a imóveis especiais e pontos estratégicos. Nesse último indicador, ressalta-se a sua importância para o combate à dengue, visto que são locais com alto potencial de disseminação do vírus e dispersão

do mosquito transmissor da doença. Esse baixo índice do Grupo 2 em relação aos demais grupos pode ser influenciado pelo número bem abaixo de agentes comunitários de saúde, quando comparado com os demais grupos.

Tabela 17 - Valores médios para os indicadores do componente Processo.

Grupo	Visita casa a casa	Controle de Criadouros	ADL	Visitas IEs/PEs
0	0,54	0,18	0,14	1,05
1	0,43	0,13	0,16	0,86
2	0,64	0,10	0,49	0,70
3	0,22	0,04	0,13	0,78

No entanto, como índice de visitas a domicílio é maior do que os demais, esses números podem estar relacionados a estratégia de ação dos municípios desse grupo. Outra hipótese pode estar relacionada a forma como essas ações estão relacionadas, ou seja, se são realizadas preventivamente ou de forma reativa, como nos cenários de alta incidência de casos de dengue. O número muito acima da avaliação de densidade larvária (ADL) pode ser outro indício da adoção de estratégia diferente nos municípios desse grupo.

Tabela 18 - Valores médios para os indicadores do componente Resultado.

Grupo	Incidência (por 100.000 hab)	Incidência por porte populacional	Infestação	Semanas com casos
0	1.225,29	369,90	1,36	17,43
1	643,03	143,91	1,32	8,42
2	498,71	291,57	1,04	26,56
3	179,91	52,23	0,79	8,20

Quando são analisados os valores dos indicadores do componente Resultado (Tabela 18), pode-se observar que os municípios do Grupo 2 têm, em média, um número maior de semanas com casos de dengue. Enquanto nos grupos 1 e 3 existem

cerca de 8 semanas em média com casos de dengue, no Grupo 2 esse número é mais do que o triplo. Ou seja, para os municípios do Grupo 2, em metade do ano há casos de dengue.

Essa grande diferença chama a atenção especialmente quando são relacionados os indicadores de semana de casos com índices de infestação. Enquanto o Grupo 1 tem menos que um terço das semanas com casos do que o Grupo 2, o índice de infestação no Grupo 1 é cerca de 31% maior do que o Grupo 2. Duas hipóteses poderiam ser verificadas a partir dessa informação: 1) o número de semanas é menor no Grupo 1 porque identifica adequadamente os pontos de infestação e toma as medidas necessárias, ou 2) as atividades de geração desse índice no Grupo 2 não estão sendo feitas adequadamente.

Em relação à incidência, observa-se correlação positiva entre a taxa de incidência e o índice de infestação. No entanto, dado que a comparação do valor da incidência pode ser influenciada pelo tamanho das populações, as diretrizes do Secretaria de Saúde do Estado de São Paulo usam um coeficiente de incidência de acordo com o porte populacional. Para esse indicador, nota-se que o Grupo 2 apresenta piores resultados que o Grupo 1 e esse, por sua vez, tem pior desempenho médio do que os municípios do Grupo 1, mesmo tendo número similares de semanas com casos.

A análise dos indicadores em conjunto mostra-se uma melhor estratégia do que a análise individual. Por exemplo, a literatura existente aponta limitações quanto ao uso do índice de infestação para avaliação da situação da dengue (BOWMAN; DONEGAN; MCCALL, 2016). Isso é corroborado pelos dados apresentados para os quatro grupos. A análise exclusiva do índice de infestação como forma de medir os resultados pode mascarar eventuais problemas de gestão e execução das atividades. De acordo com dados da Tabela 18, o índice de infestação para os Grupos 2 e 3 poderia ser considerado satisfatório ($< 1,1$). No entanto, a análise dos demais indicadores aponta resultados ruins nos outros indicadores para o Grupo 2, especialmente o número de semanas de casos e a incidência em relação ao porte populacional.

Essa aparente contradição dos indicadores pode, na verdade, revelar problemas na gestão do programa de combate à dengue. Por exemplo, de acordo com normas do

Ministério da Saúde (MS), um agente comunitário de saúde (ACS) deve ser responsável por no máximo 750 habitantes, o que seria equivalente a 1,33 agentes para cada mil habitantes. Os municípios do Grupo 2 têm, em média, quase 52% menos ACS do que o necessário preconizado pelo MS. Não se pode afirmar, no entanto, que apenas essa variável pode explicar a diferença nos resultados entre os grupos. Quando a análise inclui o Contexto, outras hipóteses podem ser apresentadas.

Na Tabela 19, são apresentados valores de um indicador de Resultado, um de Estrutura e quatro relacionados a Contexto. Nessa tabela, pode-se observar que o Grupo 2 tem maiores valores de PIB per capita e população urbana. Existe extensa literatura relacionando a questão da urbanização com os casos de dengue e talvez o alto índice de população urbana possa ser uma explicação para o pior desempenho do Grupo 2 em relação aos municípios dos grupos 1 e 3, quanto à incidência por porte e a todos os outros, em relação ao número de semanas com casos (Tabela 18). No entanto, chama a atenção que os municípios desse grupo têm, em média, o segundo pior nível de investimento de recursos próprios em saúde entre os quatro grupos, mesmo tendo o Grupo 2 a maior média de PIB per capita. Assim, dado que tem um dos fatores de risco bem maior do que os demais (população urbana), o menor investimento em saúde e o número de ACS muito abaixo do que o recomendado, podem ser esses os fatores que ajudem a explicar os resultados do Grupo 2.

Tabela 19 - Valores médios para os indicadores incluindo o contexto.

Grupo	Incidência por porte	ACS	PIB per capita	Desp Rec Próprios/hab	População Urbana (%)	Temperatura Máxima
0	369,90	1,56	26216,17	702,80	90	29,27
1	143,91	1,69	21229,74	727,12	73	27,38
2	291,57	0,69	39021,50	654,56	98	25,51
3	52,23	1,60	17711,29	567,77	64	25,32

Quanto ao Grupo 0, grupo de municípios com maior valor médio de incidência, uma possível explicação pode estar relacionada ao alto índice de população urbana e a maior temperatura máxima. No entanto, mesmo com esses fatores de risco maior do

que o Grupo 1, os municípios do Grupo 0 têm os mesmos problemas dos municípios do grupo 2: menor investimento de recursos próprios e menor número de ACS (quando comparados com o Grupo 1). Mesmo que o número de ACS seja acima do recomendado pelo MS, uma hipótese é que, dadas as condições mais favoráveis à disseminação da dengue, apenas cumprir a meta preconizada para estrutura não seja suficiente para o combate à dengue.

Outro ponto que se pode observar é que os dois piores grupos quando ao indicador incidência por porte populacional são os de maior PIB per capita. As pesquisas sobre Saúde Planetária discutem, entre outras coisas, esse paradoxo entre riqueza/desenvolvimento e saúde da população. O desenvolvimento realizado de forma não sustentável tem seus custos associados e, talvez, a grande disseminação da dengue seja um desses.

6.4.3.4.2 Análise intragrupos

Para análise dos indicadores dentro dos grupos, os municípios foram divididos em dois subgrupos de acordo com o desempenho em relação ao indicador “Incidência por porte populacional”:

- Grupo 1 (melhor): municípios que cuja incidência não atingiu 100% do coeficiente de incidência por porte populacional;
- Grupo 2 (pior): municípios que cuja incidência ultrapassou 100% do coeficiente de incidência por porte populacional.

Em seguida, foram calculados os valores médios de cada indicador dos componentes Estrutura, Processo e Resultado, além dos quatro principais indicadores referentes a Contexto, de acordo com a Figura 31. Na Tabela 20 são apresentados os resultados, com destaque em negrito para o maior valor de cada indicador, quando a diferença for superior a 0,01.

Em geral, os municípios com pior desempenho tiveram valores mais altos de quase todos os indicadores. Em relação às variáveis de Contexto, a temperatura máxima foi mais alta em todos os subgrupos de pior desempenho. Os valores referentes à distribuição de água e população urbana, por sua vez, foram mais altos em apenas

dois desses subgrupos e sendo, nos dois casos, subgrupos dos grupos 2 e 3. No Grupo 1, observa-se que os dois indicadores tiveram resultados melhores para o subgrupo de melhor desempenho, enquanto que o Grupo 0 as médias foram idênticas para os dois indicadores. Por fim, em relação ao indicador de saneamento, pode-se observar valores mais altos em três subgrupos de pior desempenho.

Para o componente Estrutura, não foi observada diferenças para o indicador referente ao número de Unidades Básicas de Saúde por habitante, com exceção do Grupo 2, em que os municípios de melhor desempenho têm, em média, o número maior para este indicador. Para os cinco indicadores referentes a recursos humanos, foram encontrados diversos. Enquanto os valores para ACS e enfermeiros foram superiores em dois subgrupos de melhor desempenho (nos grupos 0 e 3, para ACS; nos grupos 2 e 3 para enfermeiros), o número de agentes de combate às endemias (ACE) foram superiores em três subgrupos de pior desempenho, o mesmo observado para o indicador relacionado ao número de médicos. Para o indicador relacionado ao número de equipes de saúde da família (PSF), não foram observadas diferenças nos grupos.

Tabela 20 - Valores médios considerando o desempenho quanto ao coeficiente de incidência (em negrito, o valor mais alto e com diferença superior a 0.01)

Grupo	0		1		2		3	
	Melhor	Pior	Melhor	Pior	Melhor	Pior	Melhor	Pior
Resultado								
Incidência	322,14	1412,8	200,23	1181,34	154,93	625,85	103,79	653,52
Incidência por porte	61,88	406,7	40,17	270,03	63,49	375,92	25,62	217,84
Infestação	1,24	1,38	1,24	1,41	0,75	1,14	0,74	1,11
Semanas com casos	8,38	18,96	5,3	12,2	16,76	30,18	6,85	16,6
Contexto								
Temperatura máxima	28,41	29,12	27,2	28,15	25,03	25,5	24,36	25,33
Água	0,88	0,88	0,71	0,68	0,90	0,94	0,53	0,63
Esgoto	0,89	0,91	0,69	0,69	0,80	0,83	0,41	0,48
População urbana	0,88	0,89	0,72	0,71	0,93	0,97	0,61	0,80
Estrutura								
ACS	1,8	1,52	1,57	1,82	0,67	0,69	1,63	1,43
ACE	0,33	0,39	0,22	0,28	0,09	0,12	0,1	0,1
Médicos	0,88	1,17	0,88	0,91	1,98	1,88	1,02	1,12
Enfermeiros	0,94	1,03	0,92	0,95	1,18	1,06	0,94	0,72
PSF	0,08	0,08	0,1	0,1	0,07	0,06	0,12	0,13
UBS	0,29	0,3	0,35	0,35	0,18	0,12	0,24	0,24

Despesas em saúde	971,95	942,56	957,06	962,57	845,77	925,76	792,45	799,22
Despesas com Rec. Próprios	729,19	698,35	743,49	707,22	605,43	672,73	565,51	581,81
Processo								
Casa a Casa	0,47	0,55	0,42	0,45	0,21	0,79	0,24	0,09
Controle de Criadouros	0,16	0,18	0,11	0,15	0,04	0,12	0,03	0,05
Visitas IEs/PEs	0,94	1,07	0,9	0,81	0,73	0,69	0,78	0,78
ADL	0,14	0,14	0,18	0,15	0,07	0,65	0,14	0,09

Para o componente Processo, o indicador de visitas casa a casa foi superior nos municípios de pior desempenho dos grupos 0, 1 e 2. Para o Controle de Criadouros, todos os subgrupos de pior desempenho tiveram valores superiores. Em relação às visitas a Imóveis Especiais e Pontos Estratégicos, foram observados valores superiores em dois subgrupos de municípios com melhor desempenho (nos Grupos 1 e 2), enquanto que para avaliação de densidade larvária (ADL), os valores superiores foram encontrados nos subgrupos de melhor desempenho dos Grupos 1 e 3.

Análise similar foi realizada quanto ao indicador índice de infestação, em que o critério de estratificação dentro dos grupos foi o valor considerado satisfatório (índice de infestação < 1,1). Os valores médios de cada estrato/grupo são apresentados na Tabela 21, com os valores superiores apresentados em negrito.

Tabela 21 - Valores médios considerando o desempenho quanto ao índice de infestação (em negrito, o valor mais alto e com diferença superior a 0,01).

Grupo	0		1		2		3	
	Melhor	Pior	Melhor	Pior	Melhor	Pior	Melhor	Pior
Resultado								
Incidência	1278,75	1237,16	591,61	682,4	477,59	524,51	132,02	339,53
Incidência por porte	323,96	382,37	124,42	158,83	270,48	317,34	36,81	103,65
Infestação	0,62	1,93	0,67	1,81	0,51	1,68	0,46	1,92
Semanas com casos	15,86	18,66	7,92	8,8	25,66	27,65	7,3	11,21
Contexto								
Temperatura máxima	28,97	29,06	27,58	27,66	25,19	25,59	24,36	24,92
Água	0,87	0,89	0,67	0,72	0,93	0,93	0,53	0,59
Esgoto	0,91	0,91	0,67	0,70	0,83	0,81	0,42	0,44
População urbana	0,88	0,89	0,69	0,74	0,96	0,96	0,61	0,72
Estrutura								
ACS	1,59	1,54	1,58	1,77	0,66	0,71	1,64	1,48
ACE	0,41	0,37	0,29	0,22	0,09	0,13	0,11	0,07
Médicos	0,99	1,23	1,01	0,80	1,88	1,94	0,89	1,53

Enfermeiros	0,95	1,07	1,00	0,88	1,05	1,14	0,78	1,37
PSF	0,09	0,08	0,09	0,11	0,06	0,06	0,12	0,14
UBS	0,31	0,28	0,36	0,35	0,14	0,13	0,23	0,27
Despesas em saúde	1006,09	900,98	1012,00	919,39	887,74	924,23	803,18	760,75
Despesas com Rec. Próprios	760,80	657,97	768,56	695,39	644,08	667,37	570,78	557,73
Processo								
Casa a Casa	0,56	0,52	0,47	0,41	0,99	0,21	0,23	0,17
Controle de Criadouros	0,19	0,17	0,14	0,11	0,13	0,06	0,03	0,05
Visitas IEs/PEs	1,06	1,04	0,90	0,82	0,68	0,72	0,77	0,80
ADL	0,17	0,11	0,19	0,14	0,85	0,06	0,13	0,11

Assim como na análise por incidência, o indicador de temperatura máxima foi superior em todos os estratos com piores desempenhos nos grupos. No entanto, para os Grupos 0 e 2, não foram observadas grandes diferenças nos demais indicadores de contexto, ao contrário dos grupos 1 e 3, em que os indicadores de Contexto tiveram valores superiores nos municípios de pior desempenho.

Para o componente Estrutura, o indicador referente a equipes de saúde da família teve valores superiores nos substratos de pior desempenho para os Grupos 1 e 3, e valor superior para os municípios de melhor desempenho do Grupo 0. Em relação às unidades básicas de saúde, observa-se, com exceção do Grupo 3, valores mais altos nos estratos de melhor desempenho para infestação. Esse fato poderia sugerir uma relação na estrutura da atenção básica quanto ao desempenho dos municípios nas ações relacionadas ao controle de infestação. Em relação aos dois indicadores referentes às despesas com saúde, parece existir um padrão claro entre os resultados e os investimentos em saúde. Com exceção do Grupo 2, os municípios de melhor desempenho nos demais grupos foram aqueles que investiram mais recursos em saúde. Em relação aos demais indicadores de estrutura, observou-se padrões nos Grupos 1 e 2. No Grupo 1, os municípios com índice de infestação satisfatório foram aqueles com maiores valores de indicadores de recursos humanos, exceto para ACS. Enquanto que para o Grupo 2 foi observado o inverso: os municípios com maiores valores foram os de pior desempenho para infestação.

Em relação ao componente Processo, os municípios de melhor desempenho do Grupo 0 foram os que tiveram maiores médias para todos os indicadores desse

componente, o mesmo sendo observado para os municípios do Grupo 1 e Grupo 2, com exceção das visitas a IEs/PEs para esse último. A partir desses resultados, observa-se um padrão para as atividades de combate à dengue e o desempenho quanto a infestação: os municípios de melhor desempenho foram os que realizaram mais ações em média, independente do contexto (clima, saneamento, população urbana).

Outro fator a ser destacado é a relação entre número de agentes comunitários de saúde e ações realizadas. Os municípios dos estratos de pior desempenho nos Grupos 1 e 2 realizaram menos ações mesmo tendo um número médio de agentes maior do que os municípios dos estratos de melhor desempenho. Embora esse ponto não possa explicar exclusivamente os resultados para o índice de infestação, o fato de o maior número de recursos humanos não implicar em maior número de ações pode ser um indício de problemas na gestão do programa nesses locais e pesquisas futuras poderiam investigar essa hipótese.

6.5 Experimento 2: Painel BI para avaliações baseadas no modelo de Donabedian

A seguir são descritos os detalhes para o segundo experimento realizado para avaliação da metodologia proposta. Nesse experimento, é demonstrado um outro tipo de avaliação em saúde possível a partir do repositório comum de dados: monitoramento dos indicadores de forma contínua.

6.5.1 Descrição experimento

6.5.1.1 Objetivo

Criar um painel para monitoramento contínuo dos indicadores conforme o modelo de Donabedian.

6.5.1.2 Etapas

O experimento foi desenvolvido em quatro etapas, desde a seleção de variáveis até a construção do painel de indicadores para visualização dos resultados.

6.5.2 Resultados

6.5.2.1 Etapa 1: Seleção dos indicadores

Na primeira etapa, foram definidos os indicadores necessários para monitoramento, assim como os níveis de agregação espacial e temporal. Na Tabela 22 são listados os níveis de agregação para as escalas temporal e espacial, que funcionarão como filtros no painel. Dessa forma, os gestores poderão visualizar os indicadores de forma agregada ou detalhada. Embora a semana possa ser um nível desagregado do mês, na prática o mesmo não ocorre com as semanas epidemiológicas, que podem ter início em um mês e final em outro. Assim, nesse experimento, considerou-se a semana epidemiológica e o mês como níveis detalhados de ano.

Tabela 22 - Níveis de agregação utilizados como filtros no painel desenvolvido.

Nome	Tipo
Município	Espacial
Departamento Regional de Saúde (DRS)	Espacial
Ano	Temporal
Mês	Temporal
Semana epidemiológica	Temporal

Assim, de acordo com o modelo de avaliação definido na Seção 6.3.1, os indicadores selecionados para compor o painel são listados na Tabela 23. Enquanto alguns indicadores, como os do componente Estrutura, são calculados anualmente, outros foram avaliados de acordo com a semana epidemiológica, como as ações realizadas e o número de casos. Optou-se pela escala anual para os indicadores de estrutura devido a baixa variabilidade ao longo do ano. Em relação aos indicadores relacionados a número de casos e ações realizados, optou-se pelo uso da semana epidemiológica e mês para avaliar possíveis efeitos das ações dos períodos seguintes.

Tabela 23 - Indicadores selecionados para o painel Donabedian.

Indicador	Descrição	Menor nível de detalhamento
Estrutura		
Agentes comunitários de saúde (ACS)	Número de ACS por mil habitantes	Ano
Agentes de combate a endemias (ACE)	Número de ACE por mil habitantes	Ano
Médicos (MED)	Número de médicos por mil habitantes	Ano
Enfermeiros (ENF)	Número de enfermeiros por mil habitantes	Ano
Equipes de saúde da família (PSF)	Número de equipes de saúde da família por mil habitantes	Ano
Unidades básicas de saúde (UBS)	Número de UBS por mil habitantes	Ano
Despesas em saúde (DESP)	Investimento total em saúde por habitante	Ano
Despesas com Recursos Próprios (DESPPROP)	Valor investido em saúde com recursos próprios por habitante	Ano
Processo		
Visita casa a casa (CASA)	Número de visitas casa a casa por quarteirão	Semana epidemiológica e mês
Controle de criadouros (CRIA)	Número de ações de controle de criadouros realizada por quarteirão	Semana epidemiológica e mês
Visitas a Imóveis Especiais e Pontos Estratégicos (IES)	Número de ações de visitas a PEs e IEs por imóveis desse tipo	Semana epidemiológica e mês

Avaliação de densidade larvária (ADL)	Número total de ações de ADL	Semana epidemiológica e mês
Resultado		
Incidência (INCID)	Número de casos de dengue por 100000 habitantes	Semana epidemiológica e mês
Incidência por porte populacional (INCID_PORT)	Taxa de incidência em função do coeficiente de incidência correspondente ao porte populacional	Ano
Semanas com casos (SEs)	Total de semanas epidemiológicas com casos de dengue	Mês
Infestação (INF)	Índice de infestação predial	Semana epidemiológica e mês

6.5.2.2 Etapa 2: Criação do *Data Warehouse*

Em seguida, na segunda etapa, foi construído o *Data Warehouse* (DW) a ser utilizado pelo painel com o uso da plataforma em nuvem Google BigQuery. Embora parte dos dados, como ações, estivessem disponíveis no nível de granularidade diário, optou-se, para esse experimento, pelo uso da semana como menor nível de granularidade. Dependendo do objetivo de uso do painel, o uso de dados diários pode ser adotado, o que implica em: 1) Maior espaço utilizado para armazenamento no DW, e 2) Maior número de consultas possíveis.

Dessa forma, o DW foi desenvolvido segundo o esquema estrela, em que uma tabela de fatos é relacionada a diversas tabelas de dimensões. As dimensões são as tabelas que contém informações gerais que não precisam se repetir, como o nome do município e o departamento regional a que pertence. Já as tabelas de fatos contém os dados históricos, que podem variar ao longo do tempo.

No caso desse experimento, para evitar redundâncias, foram criadas três tabelas de fatos correspondentes aos três componentes do modelo de Donabedian. Os indicadores referentes à Estrutura foram associados às dimensões Município, DRS e Data, conforme diagrama na Figura 32. Nessa figura, as tabelas com prefixo *DIM* correspondem às tabelas de dimensões, enquanto que o prefixo *FACT* é referente às tabelas de fatos. A tabela *Pop* contém o total da população da cidade no período e é utilizada para cálculo das informações em função do total de habitantes do município e, com isso, possibilitar comparações entre os locais.

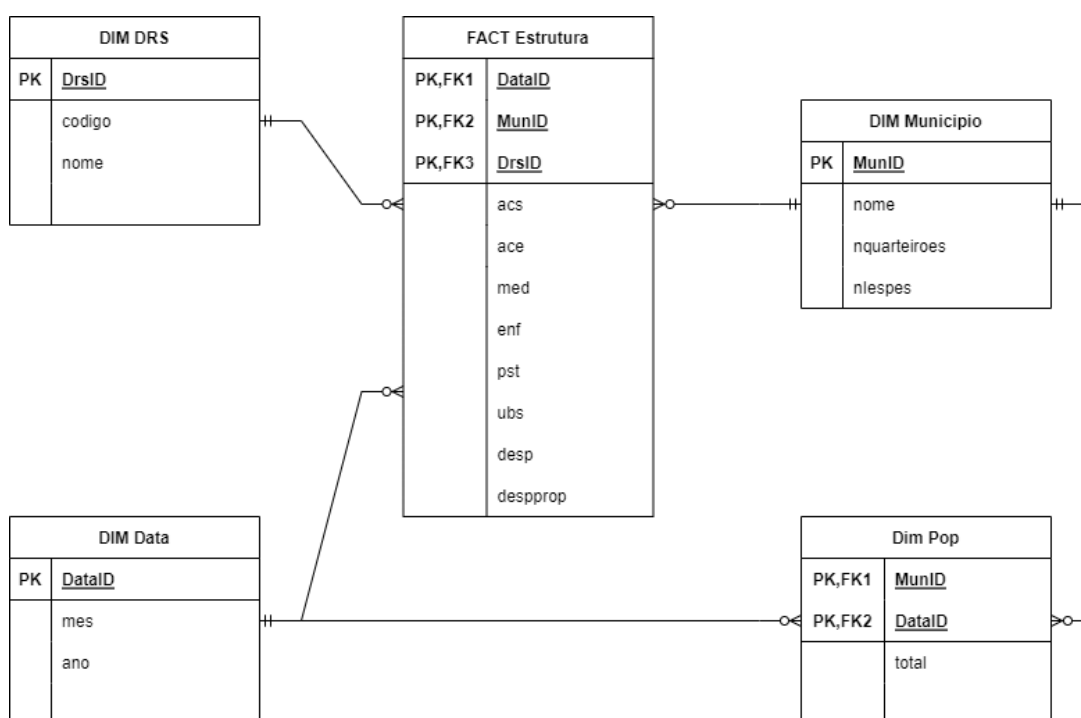


Figura 32 - Esquema do Data Warehouse para os indicadores de Estrutura.

Em relação aos quatro indicadores do componente Processo, além das dimensões descritas anteriormente, foi incluída a dimensão referente à semana epidemiológica no esquema pois, de acordo com o documento de diretrizes de combate à dengue (SECRETARIA DE SAÚDE DE SÃO PAULO, 2017), são realizadas análises da situação de risco para dengue em função da semana epidemiológica. O diagrama correspondente ao esquema estrela para os indicadores de processo é apresentado na Figura 33.

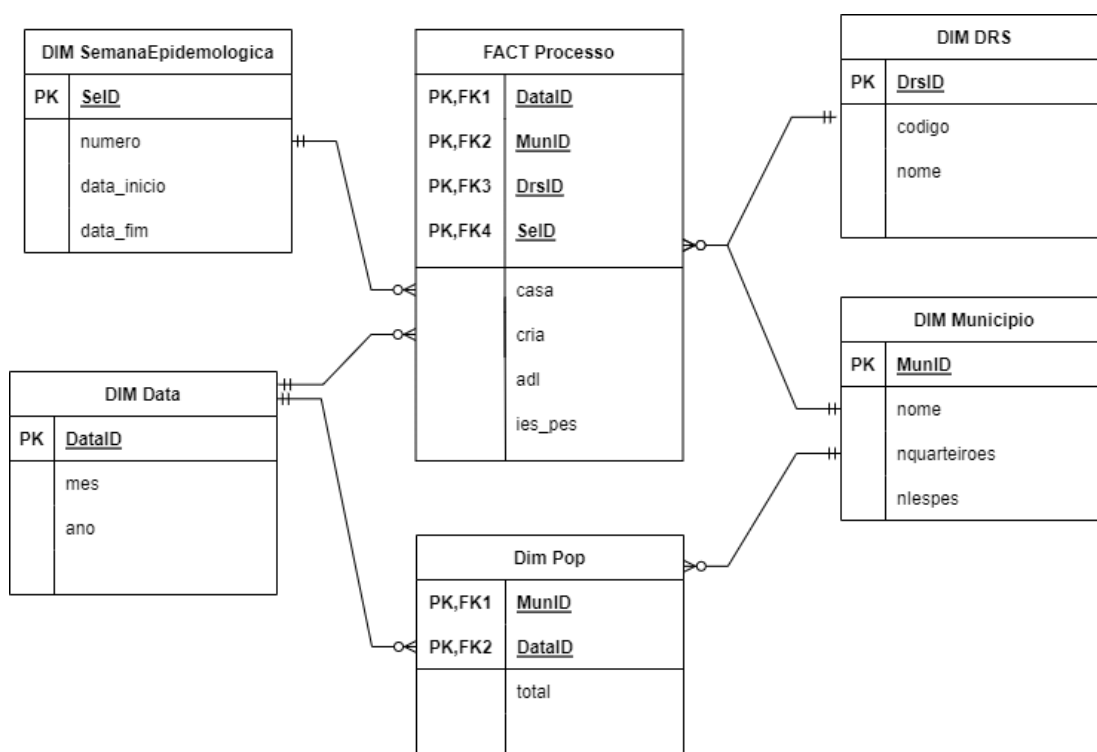


Figura 33 - Esquema do Data Warehouse para os indicadores de Processo.

Por fim, foi incluída a tabela de fatos referente ao componente Resultado, conforme Figura 34. Ao contrário dos outros dois componentes, a tabela de fatos de Resultado incluiu apenas o número de casos e o índice de infestação do período e local. A razão para isso é que os três indicadores resultados serão calculados no desenvolvimento do painel: a incidência, número de semanas com casos e incidência de acordo com o porte populacional.

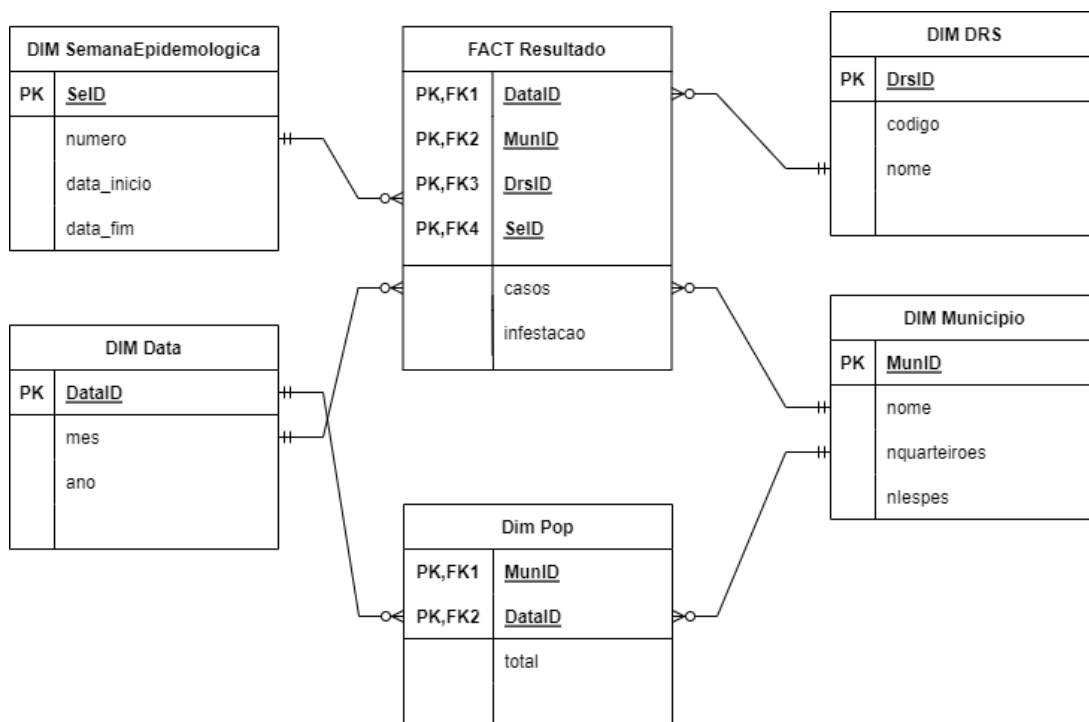


Figura 34 - Esquema do Data Warehouse para os indicadores de Resultado.

6.5.2.3 Etapa 3: Aquisição dos dados

Após o desenvolvimento da estrutura do *Data Warehouse*, para que a avaliação seja realizada de forma contínua, foram desenvolvidos *scripts* em Python para ingestão periódica dos dados a partir de estrutura de dados desenvolvida. Esses *scripts*, coletam os dados em suas formas originais, realizam diversas transformações e fazem a ingestão no DW. As transformações nos dados originais foram as mesmas realizadas no Experimento 1.

6.5.2.4 Etapa 4: Criação do Painel

O painel foi desenvolvido usando a ferramenta Google Data Studio, que permite acesso controlado em nuvem, não necessitando de instalação de pacotes de software, já que o painel pode ser acessado via navegador Web.

Esse ambiente permite a criação de diversas formas de visualização dos dados, que podem ser organizados por assunto. No painel desenvolvido para este experimento, a visualização inicial (Figura 35) apresenta três indicadores: índice de

infestação, total de número de casos e casos por semana epidemiológica. Para os dois primeiros foram apresentados os dados de 2019 e o desempenho em relação ao ano anterior, enquanto no terceiro é apresentada a série histórica.

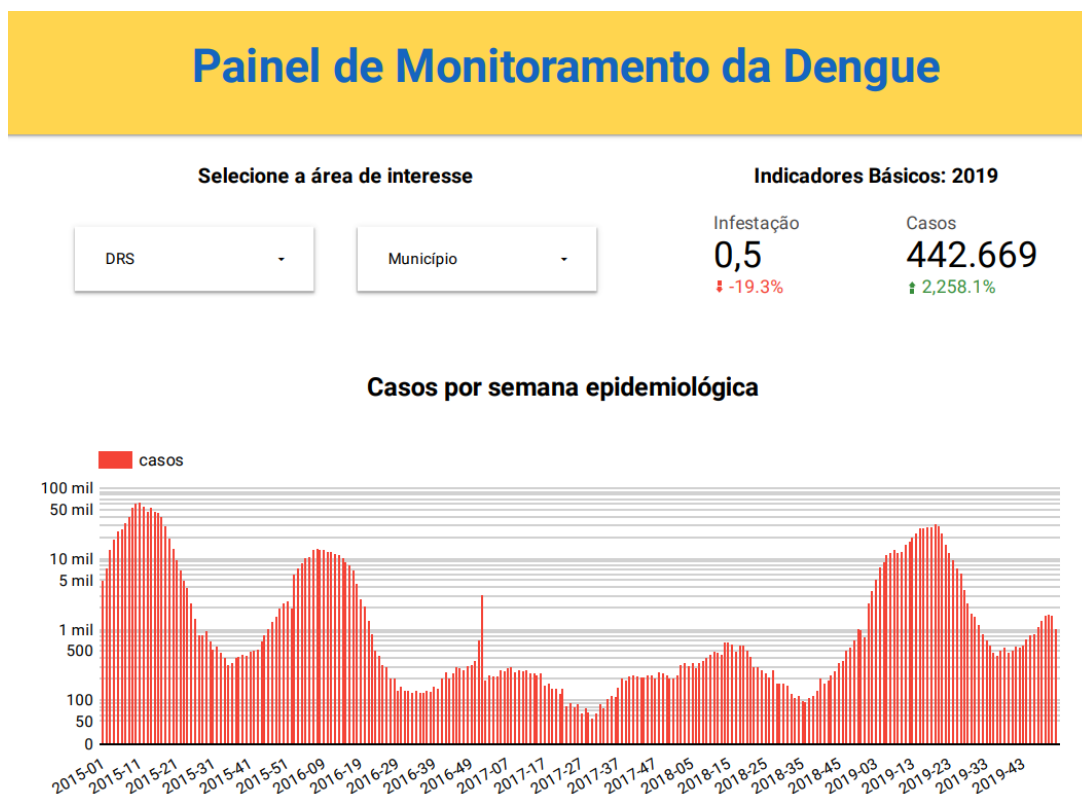


Figura 35 - Apresentação inicial do painel de monitoramento da dengue.

De modo a permitir análise detalhada pelos gestores, foram disponibilizados dois filtros nessa página inicial do painel: Departamento Regional de Saúde (DRS) e Município. Dessa forma, a análise dos indicadores pode ser realizada de acordo com a região de interesse. Por exemplo, para a DRS 01, que representa a Grande São Paulo, nota-se um aumento do índice de infestação, ao contrário do observado pela média geral do estado (Figura 36).

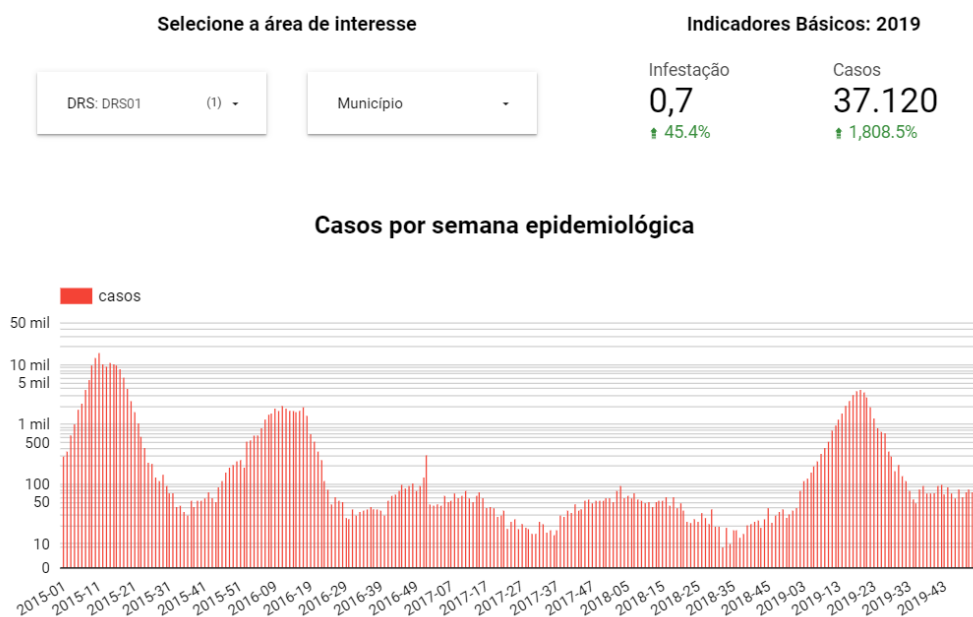


Figura 36 - Dados visualizados para a DRS 01.

Em seguida, foram desenvolvidos painéis para cada componente do modelo de Donabedian. No primeiro painel, o desempenho para os indicadores do componente Resultado é apresentado em uma espécie de medidor para metas. Para essa forma de visualização dos dados, pode-se estabelecer níveis de classificação, em que cada nível está associado a um intervalo de valores. Para os indicadores de Resultado, os níveis definidos em conjunto com os especialistas e presentes em normas são detalhados na Tabela 24. Através dessa forma de visualização, os gestores podem rapidamente identificar possíveis problemas nos resultados do programa e estabelecer medidas de acordo com a situação.

Tabela 24 - Níveis definidos para medição do desempenho dos indicadores de Resultado.

Indicador	Nível Bom	Nível de Atenção	Nível Ruim
Incidência	0 - 20	20 – 50	Maior que 50
Infestação	0 – 1	1 – 3,9	Maior que 3,9
Incidência por porte populacional	0 – 20	20 – 50	Maior que 50
Semanas com casos	0 – 10	10 – 20	Maior que 20

De acordo com os resultados apresentados na Figura 37, observa-se que a meta foi atingida apenas para o indicador infestação. A possível contradição desse indicador, cuja medição não é obrigatória, com os indicadores relacionados ao número de casos, de notificação compulsória, podem ser um indício de possíveis problemas no uso isolado dos níveis de infestação como forma de avaliar o sucesso do programa de combate à dengue. Assim como nos demais painéis, a utilização de filtros permite a visualização dos resultados de acordo com a região ou período.



Figura 37 - Indicadores do componente Resultado.

Os painéis podem atender a públicos com diversos interesses, como as pessoas responsáveis pela gestão das atividades, assim como aquelas com objetivos de desenvolvimento de pesquisas. Dessa forma, além dos elementos para visualização rápida dos resultados, os dados podem ser apresentados em formas tradicionais, como tabelas. No entanto, mesmo essas formas tradicionais podem fazer uso de recursos que tornem sua utilização mais interativa, como o uso de filtros e ordenação dos registros quanto a algum parâmetro, conforme Figura 38. A ferramenta adotada para o painel desenvolvido para este experimento apresenta uma opção de *download* dos dados, de maneira que outras formas de análises possam ser aplicadas aos dados selecionados, como análise estatísticas e algoritmos de *machine learning*.

Selecione os filtros desejados

Município: DRS: Ano: 2019 (1)

	nome	ano	Incidência (*)	Incidência por porte (**)	Infestação	Semanas com casos
1.	São Joaquim da Barra	2019	12.536,62	4.178,87	1,32	42
2.	Potim	2019	11.540,80	3.846,93	2,14	45
3.	Bilac	2019	10.057,26	1.676,21	0,00	26
4.	Araraquara	2019	9.880,88	6.587,25	0,66	53
5.	Nova Aliança	2019	9.809,26	1.634,88	0,41	39
6.	Tupã	2019	9.675,84	3.225,28	0,38	47
7.	Jaborandi	2019	9.092,22	1.515,37	0,18	25
8.	Paulo de Faria	2019	9.066,52	1.511,09	1,21	34
9.	Bady Bassitt	2019	9.004,69	3.001,56	0,22	43
10.	Andradina	2019	8.527,39	2.842,46	0,63	53

Figura 38 - Visualização detalhada dos indicadores do componente Resultado.

Em seguida, foram desenvolvidos painéis para os indicadores referentes ao componente Estrutura. Para esses indicadores, foram criados elementos para visualização dos indicadores, como o medidor baseado em níveis, assim como a visualização dos valores considerando a série temporal. Na Figura 39, são apresentadas visualizações de três indicadores desse componente: número de Agentes Comunitários de Saúde (ACS) por habitante e os dois indicadores relacionados a despesas em saúde por habitante. Utilizando o exemplo da DRS 01, nota-se que, apesar do aumento de investimento ao longo do tempo, o número de ACS foi abaixo do recomendado pelo Ministério da Saúde. Esse fato pode ser uma das razões para os valores apresentados anteriormente para essa região nos indicadores do componente Resultado.

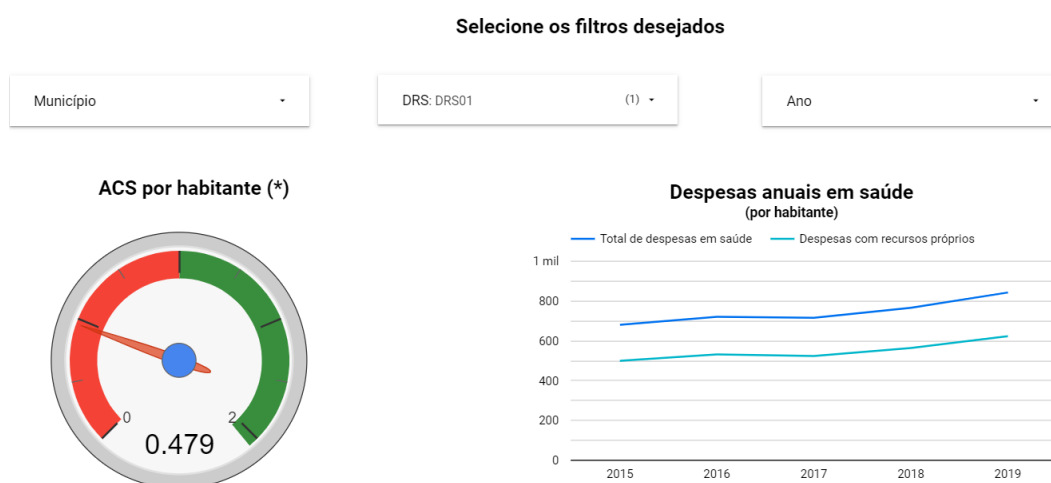


Figura 39 - Indicadores do componente Estrutura.

Assim como o componente Resultado, também foi desenvolvido um painel para visualização detalhada dos indicadores do componente Estrutura, conforme Figura 40. Através dessa forma de visualização interativa dos dados, pode-se realizar análises detalhadas de quais locais apresentam resultados abaixo da meta para esses indicadores. Através da Figura 40 é possível observar quais municípios apresentam número de ACS abaixo do recomendado. Com isso, os órgãos de gestão e controle podem realizar políticas para melhoria desses fatores e monitoramento do uso dos recursos públicos destinados à saúde quanto ao que é preconizado pelo Ministério da Saúde e Secretaria Estadual de Saúde. Outra função importante para essa análise detalhada é possibilitar a investigação de possíveis problemas nos dados, que podem indicar deficiências tanto nos órgãos responsáveis pela notificação de informações assim como os sistemas de informação existentes.

Município ▾

DRS: DRS01 (1) ▾

Ano: 2019 (1) ▾

	nome	ACS ▾	ACE	Enf	Med	PSF	UBS	Desp	Desp rec. próp.
1.	Guararema	0	0,27	2,32	1,21	0,00	0,10	1.433,11	981,92
2.	Salesópolis	0	0,00	0,47	0,53	0,00	0,12	649,74	456,93
3.	Rio Grande da Serra	0,01	0,00	0,31	0,39	0,12	0,16	406,42	218,01
4.	Carapicuíba	0,05	0,01	1,52	0,96	0,00	0,04	345,20	253,89
5.	Cotia	0,07	0,01	1,75	1,24	0,04	0,10	603,09	452,59
6.	Caieiras	0,09	0,12	3,26	1,41	0,01	0,12	630,35	598,04
7.	Barueri	0,1	0,16	5,79	2,01	0,00	0,07	3.066,22	2.783,32
8.	Mogi das Cruzes	0,18	0,08	2,61	1,73	0,01	0,08	724,71	416,24
9.	Santo André	0,23	0,01	4,16	2,19	0,05	0,04	787,84	595,30
10.	Embu das Artes	0,27	0,08	0,98	0,56	0,05	0,06	678,52	400,35

Figura 40 - Visualização detalhada dos indicadores do componente Estrutura.

Por fim, foi desenvolvido um painel para a visualização dos indicadores correspondentes ao componente Processo (Figura 41). Para esses indicadores, os resultados são apresentados em forma de gráfico de barras com detalhamento mensal, pois as ações realizadas dependem do período do ano. Assim, para cada mês, são apresentados os valores em função do número de bairros, imóveis especiais e pontos estratégicos.



Figura 41 - Indicadores do componente Processo.

Para a DRS 01 no ano de 2019, observa-se um aumento gradual nas ações de controle de criadouros entre os meses de janeiro e junho, enquanto as ações de visitas casa-a-casa têm o maior aumento entre os meses de setembro e novembro daquele ano. A visualização da relação desses indicadores em função do período do ano pode ser útil para avaliar se a estratégia das ações está sendo realizada de forma adequada. O mesmo se aplica as ações de visitas a imóveis especiais e pontos estratégicos que, para o ano de 2019, mostraram-se com pouca variabilidade para os municípios da DRS 01.

Outra forma de avaliação das estratégias adotadas é a visualização da série histórica em função dos resultados. Na Figura 42, por exemplo, são apresentados os dados históricos por semana epidemiológica de dois indicadores de Processo (visitas casa a casa e controle de criadouros) em relação ao índice de infestação. Esse tipo de visualização pode possibilitar a análise das estratégias quanto a questões temporais. Uma questão que poderia ser respondida com esse tipo de análise é se as ações de visitas a casa a casa e controle e criadouros foram realizadas com no período adequado para controle dos índices de infestação.

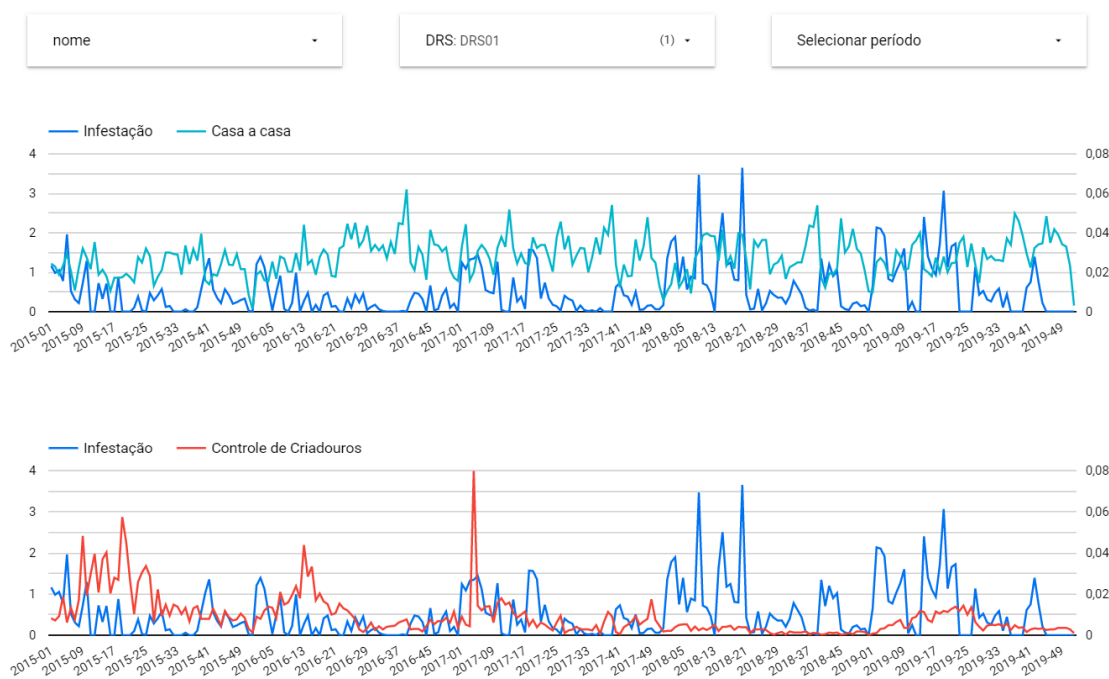


Figura 42 - Série histórica do índice de infestação em relação a indicadores do componente Processo.

No painel também foram adicionados elementos para visualização espacial dos indicadores, que pode ser uma forma útil de avaliar possíveis relações do contexto local com os resultados. Na Figura 43, são apresentados os valores do indicador Semanas com casos para os municípios de São Paulo. Esse tipo de visualização pode prover informações rápidas quanto aos resultados. Por exemplo, observa-se, na Figura 43, que os municípios com maior frequência de casos de dengue são da Grande São Paulo, região Norte e litoral norte do estado, principalmente.

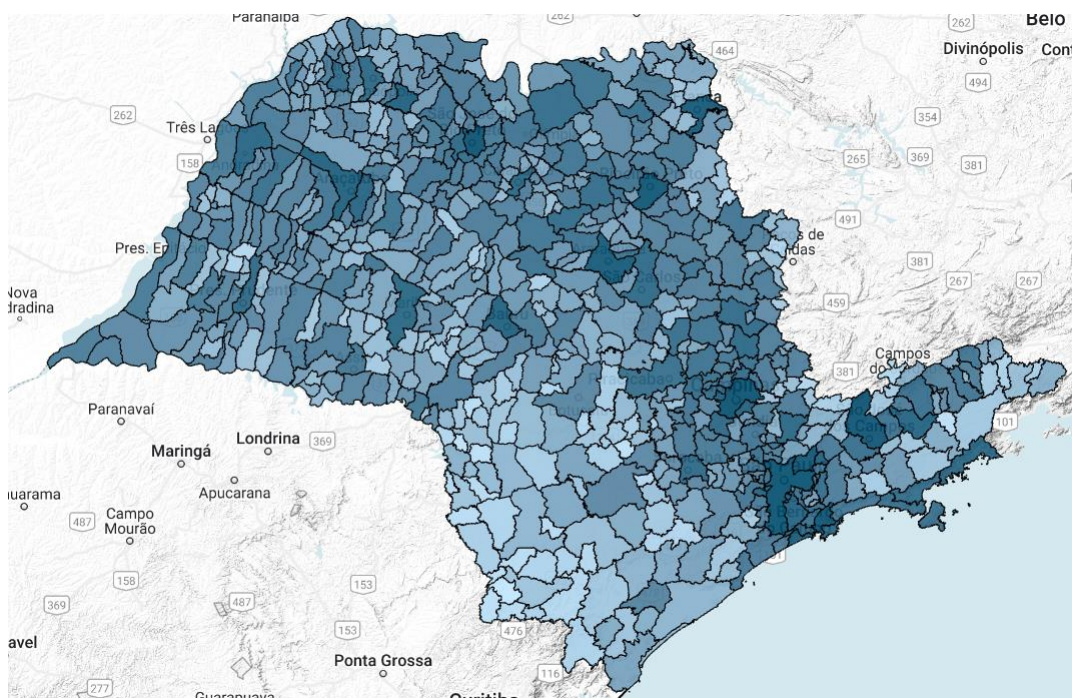


Figura 43 - Visualização espacial do indicador Semana com casos (média de 2015 a 2019).

O mesmo tipo de visualização foi desenvolvido para o indicador referente ao índice de infestação para o ano de 2019. Com base em quatro categorias, listadas na Tabela 25, os municípios foram classificados quanto ao grau de infestação. Observa-se, na Figura 44, que a maior parte das localidades com mais semanas com casos tiveram índice baixo de infestação. Esse tipo de resultado pode evidenciar deficiências nas atividades de medição desse índice.

Tabela 25 - Categorias para classificação do índice de infestação.

Categoria	Intervalo
1	0 – 1,1
2	1,1 – 3,9
3	Maior que 3,9

7 CONSIDERAÇÕES FINAIS

Este trabalho de pesquisa teve como objetivo o desenvolvimento de uma metodologia baseada em Ciência de Dados para avaliação do programa de combate à dengue. Tendo como referencial teórico o modelo de avaliação em saúde de Avedis Donabedian, esta pesquisa buscou agregar técnicas e métodos de Ciência de Dados para adequar essa atividade no cenário de *Big Data*. A avaliação em saúde é uma atividade fundamental para o monitoramento de políticas públicas e a maior disponibilidade de fontes de dados traz consigo oportunidades e desafios na área.

Assim como outras condições de saúde, a dengue é relacionada a diversos fatores. Dessa forma, não existem respostas simples para explicar a razão pela qual os casos da doença continuam a aumentar a despeito das ações executadas. Podem existir muitos motivos que podem variar, inclusive, de acordo com o local avaliado. O modelo proposto por Avedis Donabedian, amplamente utilizada na saúde, inclui três dimensões para avaliação das condições de saúde: estrutura, processo e resultado. No entanto, todos esses fatores também estão relacionados ao contexto em que estão inseridos.

As avaliações em saúde, conforme demonstrado na avaliação dos trabalhos relacionados, normalmente utilizam-se de formas de coletas de dados de alto custo, que demandam grande esforço e tempo de quem realiza a pesquisa. Não se questiona a validade desses instrumentos, muito pelo contrário. No entanto, a aplicação desses modelos de avaliação tem limitações quanto a escalabilidade e velocidade. O método apresentado nesta pesquisa de doutorado permite a execução de avaliações em áreas maiores e com maior velocidade para extração das informações. Ressalta-se, no entanto, que o método proposto não substitui as formas comumente adotadas para avaliação, mas pode ser tratada como uma forma complementar e integradora das atividades de avaliação realizadas localmente.

A pandemia da COVID-19 mostrou como a integração dos locais e velocidade representam importante fatores na disseminação da doença. Assim, a avaliação da situação local pode considerar fatores além dos seus limites geográficos. Ademais, identificar rapidamente as condições de saúde pode ser fundamental para a definição

das medidas adequadas para conter o surto de uma doença ou, ainda, atuar ativamente antes de que o surto aconteça. Com a metodologia proposta nesta pesquisa, as avaliações podem adotar métodos para integração das variáveis de contexto para auxiliar na compreensão porque determinadas ações têm mais sucesso em alguns locais em relação a outros.

Ao mesmo tempo, a crescente informatização das atividades gera um ativo de grande valor para avaliação das políticas: dados. A estruturação do Sistema Único de Saúde no Brasil faz com que todos os entes federativos tenham atuação em diferentes áreas e, com isso, geram um grande volume de dados potencialmente útil. No entanto, essa descentralização de atividades também apresenta alguns desafios para uso desses dados, como diversidade de sistemas informatizados, recursos humanos e materiais inadequados e muitos outros. Com isso, embora existam muitos dados sobre as atividades, nem sempre eles são facilmente acessíveis. A arquitetura de dados proposta nesta pesquisa pode ser adotada para facilitar a integração de dados de diversas fontes.

Nesse sentido, a principal contribuição desse trabalho de pesquisa é integrar o conhecimento estabelecido de avaliações em saúde a esse cenário de *Big Data*. Através de técnicas de Ciência de Dados, a metodologia proposta neste trabalho busca minimizar duas limitações dos métodos atualmente utilizados para avaliação em saúde: escalabilidade e velocidade. Com a integração dos dados dos múltiplos entes federativos, é possível avaliar as condições de saúde não apenas em um local, mas também em relação a seus vizinhos. Além disso, através da integração contínua das fontes de dados existentes, as avaliações quanto aos diversos indicadores do modelo de Donabedian podem ser realizadas mais rapidamente, conforme demonstrado no Experimento 2 desta pesquisa.

Outra contribuição desta pesquisa foi o método apresentado no Experimento 1. Considerando que o contexto pode influenciar nos resultados, a avaliação dos indicadores de saúde quanto aos componentes do modelo de Donabedian pode ser realizada dentro de grupos com as mesmas características que poderiam influenciar nos resultados, como temperatura e uso do solo. Foi possível observar a deficiência em alguns indicadores de estrutura, como número de agentes comunitários de saúde,

mesmo nos municípios com condições mais favoráveis para a disseminação da dengue. Esse tipo de informação pode ser útil para a gestão pública no planejamento do uso de recursos, alocando os recursos de acordo com as deficiências mais críticas em cada grupo de municípios.

Entre as possíveis limitações dos experimentos desenvolvidos nesta pesquisa, destacam-se os problemas de subnotificação de casos e vieses das medidas de infestação. A subnotificação de casos é um problema para todas as doenças e limita o planejamento e execução das ações. A análise das razões da subnotificação, como estrutura inadequada e falhas nas ações de comunicação, pode ser objeto de estudos futuros.

As medidas de infestação são objeto de críticas em diversos estudos científicos, não devendo ser usadas unicamente para avaliação em saúde. Como realizado nessa pesquisa, a análise dessas medidas pode ser feita em conjunto com outros indicadores. Além disso, a comparação entre resultados de indicadores versus incidência de casos pode ser utilizada como forma de avaliar se as próprias atividades de análise de infestação estão sendo feitas corretamente. Como a execução dessas ações são realizadas pelos municípios, a análise da qualidade das medidas de infestação pode ser realizada por órgãos de gestão e controle estaduais e federal.

Os resultados dos dois experimentos também demonstraram como uma metodologia baseada em Ciência de Dados pode agregar às avaliações em saúde. A interação com os especialistas de domínio, característica básica de pesquisas em Ciência de Dados, fez com que a pesquisa fosse orientada aos problemas definidos junto a esses especialistas. Ademais, a participação dos especialistas durante todo o ciclo de atividades possibilita que os resultados dessa pesquisa possam ser incorporados em menor tempo pela gestão do programa de combate à dengue.

Com os resultados apresentados nesta tese, pode-se concluir que a metodologia apresentada nesta pesquisa permitiu uma avaliação ampla do programa de combate à dengue no estado de São Paulo. O uso de variáveis de contexto na avaliação, como as climáticas, possibilitou a identificação de diversos pontos de melhoria, como a necessidade da contratação de mais agentes comunitários de saúde especialmente nos locais mais suscetíveis. Foi possível observar que locais com melhores resultados

tenham condições de estrutura mais adequadas. A adoção dessa metodologia de avaliação pode ser útil para a gestão pública avaliar o uso de recursos e direcionar os investimentos de forma mais coerente com as necessidades locais.

Como ressaltado anteriormente, não existem respostas simples para explicar a o avanço da dengue mesmo com as políticas públicas de controle. No entanto, com a proposta de criação de um repositório de dados a partir de fontes de dados com relações diretas e indiretas com a dengue, outras pesquisas podem ser desenvolvidas na área com objetivos além dos apresentados nesta pesquisa.

A relação de influência entre os municípios pode ser objeto de pesquisas relacionadas à dengue. Uma hipótese é que, mesmo que um município cumpra as metas em relação aos seus indicadores de estrutura e processo, os seus resultados sejam influenciados por vizinhos com baixo cumprimento das suas metas, em especial se houver grande mobilidade entre eles.

Outro possível trabalho futuro é a reavaliação das metas de indicadores conforme preconizado pelo Ministério da Saúde. A recomendação atual para agentes comunitários de saúde é de um agente para no máximo 750 habitantes. No entanto, esse número mínimo poderia considerar a suscetibilidade do município para a dengue, como temperatura, uso do solo e taxa de população urbana. Uma pesquisa poderia identificar qual seria o número ótimo de agentes de acordo com as condições locais.

A Declaração de São Paulo sobre Saúde Planetária, documento produzido e assinado por representantes de 250 organizações de 47 países, indica mensagens-chave para diversos setores (MYERS; PIVOR; SARAIVA, 2021). Para a área de tecnologia, são destacadas diversas necessidades, como desenvolvimento de ferramentas considerando demandas e características locais e o envolvimento de várias disciplinas para promoção da saúde planetária. Nesse sentido, estudos futuros poderiam apoiar-se nessas recomendações para o desenvolvimento de técnicas e mecanismos para utilização dos dados como forma de integração de diversas áreas do conhecimento para avanço no conhecimento sobre doenças. As pesquisas em Ciência de Dados podem mostrar o potencial dessa abordagem como forma de diminuição dos silos de dados e conhecimento entre as diversas áreas relacionadas à saúde planetária.

Dada a importância da diversidade das fontes de dados para a avaliação em saúde de forma sistêmica, a questão dos padrões de dados tem impacto nos processos de integração de dados de diversas áreas. Essa questão não foi o foco dessa pesquisa de doutorado e a padronização foi realizada de forma a atender os requisitos dos experimentos conduzidos. Existem iniciativas para definição de padrões em áreas específicas, como biodiversidade (GBIF, 2023), clima (UCAR, 2023) e saúde (OHDSI, 2023), e a integração desses padrões com fins de avaliação em saúde poderia ser objeto de análise de estudos futuros.

Além disso, iniciativas relacionadas a dados abertos na área da saúde, considerando as características de *Big Data*, poderiam ser objeto de estudos futuros. Existem diversas bases de dados públicas que podem ser de interesse da saúde, como as que foram utilizadas para o desenvolvimento desta pesquisa de doutorado. Estudos futuros poderiam avaliar a adequação dos sistemas públicos de dados abertos de interesse da saúde quanto aos princípios FAIR, além de questões relacionadas a Lei Geral de Proteção de Dados e outras iniciativas de dados abertos. Esses estudos poderiam auxiliar a gestão pública na evolução de suas plataformas de disponibilização de dados, possibilitando que mais pesquisas possam ser realizadas e permitindo maior integração dos dados.

Futuras pesquisas poderiam avaliar a adequação dos dados de saúde para os princípios estabelecidos em iniciativas como a GO Fair e OGP. A maior disponibilidade e interoperabilidade de dados públicos, assim como evolução das políticas de compartilhamento de dados, poderia aumentar o potencial dos resultados de avaliações em saúde como as realizadas por meio da metodologia proposta nesta pesquisa.

Finalmente, como os cenários apontam para uma piora da crise de saúde planetária, estimulada pelas mudanças climáticas, as ações para revertê-la se tornam ainda mais urgentes. Essas questões são complexas e devem ser abordadas de maneira multidisciplinar, considerando os múltiplos fatores e atores envolvidos. Nesse contexto, o aporte da computação, da ciência de dados, da inteligência artificial se torna cada vez mais fundamental. Neste trabalho ficou evidente o apoio que a tecnologia pode trazer para o entendimento e as ações de mitigação e combate a um

problema multifacetado como a dengue, e a necessidade de essas tecnologias serem incorporadas à tomada de decisão.

Referências

- ACOSTA, A. L. et al. Interfaces à transmissão e spillover do coronavírus entre florestas e cidades. **Estudos Avançados**, v. 34, p. 191–208, 2020.
- ADNAN, K.; AKBAR, R. An analytical study of information extraction from unstructured and multidimensional big data. **Journal of Big Data**, v. 6, n. 1, p. 91, 17, 2019.
- AGARWAL, R.; DHAR, V. Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. **Information Systems Research**, v. 25, n. 3, p. 443–448, 2014.
- AHMAD, F. et al. Prediction of COVID-19 Cases using Machine Learning for Effective Public Health Management. **Computers, Materials & Continua**, v. 66, n. 3, p. 2265–2282, 2021.
- ALESSANDRETTI, L. What human mobility data tell us about COVID-19 spread. **Nature Reviews Physics**, v. 4, n. 1, p. 12–13, 2021.
- ARANTES, L. J.; SHIMIZU, H. E.; MERCHÁN-HAMANN, E. Contribuições e desafios da Estratégia Saúde da Família na Atenção Primária à Saúde no Brasil: revisão da literatura. **Ciência & Saúde Coletiva**, v. 21, p. 1499–1510, 2016.
- ARAÚJO, A. B. **Uma análise das políticas de controle e combate à dengue no Brasil**. 2018. Dissertação (Mestrado em Gestão e Economia da Saúde) 157p. Universidade Federal de Pernambuco, Recife, 2018.
- ARMBRUST1, M. et al. **Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics**. In: **ANNUAL CONFERENCE ON INNOVATIVE DATA SYSTEMS RESEARCH**, 11. Anais. 2021.
- ASSEMBLÉIA LEGISLATIVA DO ESTADO DE SÃO PAULO. **Índice Paulista de Responsabilidade Social**. Disponível em: <http://www.iprs.seade.gov.br>. Acesso em: 28 fev. 2022.
- BARBOSA, G. L. et al. Indicadores de infestação larvária e influência do porte populacional na transmissão de dengue no estado de São Paulo, Brasil: um estudo ecológico no período de 2007-2008. **Epidemiologia e Serviços de Saúde**, v. 21, n. 2, p. 195–204, 2012.
- BARBOSA, G. L. et al. Influence of strategic points in the dispersion of *Aedes aegypti* in infested areas. **Revista de Saúde Pública**, v. 53, p. 1-12, 2019.
- BARMAK, D. H. et al. Dengue epidemics and human mobility. **Physical Review E**, v. 84, n. 1, p. 11901, 2011.
- BARONE, M. T. U. et al. The impact of COVID-19 on people with diabetes in Brazil. **Diabetes research and clinical practice**, v. 166, p. 108304, 2020.
- BAVIA, L. et al. Epidemiological study on dengue in southern Brazil under the perspective of climate and poverty. **Scientific Reports**, v. 10, n. 1, p. 1–16, 2020.

BEAM, A. L.; KOHANE, I. S. Big Data and Machine Learning in Health Care. **JAMA**, v. 319, n. 13, p. 1317, 2018.

BHALOTRA, S. R. et al. Productivity effects of Dengue in Brazil. **ISER Working Paper Series**, n. 2019-04, 2019.

BHATT, S. et al. The global distribution and burden of dengue. **Nature**, v. 496, n. 7446, p. 504–507, 2013.

BICALHO, D. et al. Evaluation of quality indicators for management of the National School Feeding Program in Brazil: a systematic review. **Ciência & Saúde Coletiva**, v. 26, n. 8, p. 3099–3110, 2021.

BONITA, R.; BEAGLEHOLE, R.; KJELLSTRÖM, T. **Epidemiologia básica**. Washington, D.C.: OPS, 2008.

BOWMAN, L. R.; DONEGAN, S.; MCCALL, P. J. Is Dengue Vector Control Deficient in Effectiveness or Evidence?: Systematic Review and Meta-analysis. **PLOS Neglected Tropical Diseases**, v. 10, n. 3, p. e0004551, 2016.

BRAGA, I. A.; VALLE, D. *Aedes aegypti*: histórico do controle no Brasil. **Epidemiologia e serviços de saúde**, v. 16, n. 2, p. 113–118, 2007.

BRASIL. **Lei nº 8.080**, de 19 de setembro de 1990. Dispõe sobre as condições para a promoção, proteção e recuperação da saúde, a organização e o funcionamento dos serviços correspondentes e dá outras providências. Diário Oficial da União, Brasília, DF, 1990.

BRASIL. **Lei nº 11.350**, de 05 de outubro de 2006. Regulamenta o § 5º do art. 198 da Constituição, dispõe sobre o aproveitamento de pessoal amparado pelo parágrafo único do art. 2º da Emenda Constitucional nº 51, de 14 de fevereiro de 2006, e dá outras providências. Diário Oficial da União, Brasília, DF, 2006.

BRASIL. **Portaria nº 1.378**, de 09 de julho de 2013. Regulamenta as responsabilidades e define diretrizes para execução e financiamento das ações de Vigilância em Saúde pela União, Estados, Distrito Federal e Municípios, relativos ao Sistema Nacional de Vigilância em Saúde e Sistema Nacional de Vigilância Sanitária. Diário Oficial da União, Brasília, DF, 2013.

BRASIL. **Lei nº 13.595**, de 05 de janeiro de 2018. Altera a Lei nº 11.350, de 5 de outubro de 2006, para dispor sobre a reformulação das atribuições, a jornada e as condições de trabalho, o grau de formação profissional, os cursos de formação técnica e continuada e a indenização de transporte dos profissionais Agentes Comunitários de Saúde e Agentes de Combate às Endemias. Diário Oficial da União, Brasília, DF, 2018.

BRASIL. **Lei nº 13.709**, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Diário Oficial da União, Brasília, DF, 2018.

BRELÀZ, G.; CRANTSCHANINOV, T. I.; BELLIX, L. Open Government Partnership na cidade de São Paulo e o programa São Paulo Aberta: desafios na difusão e institucionalização de uma política global. **Cadernos EBAPE.BR**, v. 19, n. 1, p. 123–137, 2021.

BÜCKER, M. et al. Transparency, auditability, and explainability of machine learning models in credit scoring. **Journal of the Operational Research Society**, v. 73, n. 1, p. 70–90, 2022.

BURKETT-CADENA, N. D.; VITTOR, A. Y. Deforestation and vector-borne disease: forest conversion favors important mosquito vectors of human pathogens. **Basic and applied ecology**, v. 26, p. 101–110, 2018.

CAMPOS, A. C. **Duas em cada três favelas no país estão a menos de 2 km de hospitais**. Disponível em: <https://agenciabrasil.ebc.com.br/economia/noticia/2020-05/duas-em-cada-tres-favelas-estao-a-menos-de-dois-quilometros>. Acesso em: 28 jul. 2022.

CARABALI, M. et al. Why are people with dengue dying? A scoping review of determinants for dengue mortality. **BMC infectious diseases**, v. 15, n. 1, p. 1–14, 2015.

CÁRCAMO, A. et al. Informed community mobilization for dengue prevention in households with and without a regular water supply: secondary analysis from the Camino Verde trial in Nicaragua. **BMC Public Health**, v. 17, n. 1, p. 79–84, 2017.

CHAVES, L. A. et al. Programa Nacional de Avaliação dos Serviços de Saúde (PNASS) 2015-2016: uma análise sobre os hospitais no Brasil. **Revista Brasileira de Epidemiologia**, v. 24, p. e210002, 2021.

CHENG, P. et al. Predominant airborne transmission and insignificant fomite transmission of SARS-CoV-2 in a two-bus COVID-19 outbreak originating from the same pre-symptomatic index case. **Journal of hazardous materials**, v. 425, p. 128051, 2022.

CHIARAVALLOTI NETO, F. et al. Controle do dengue em uma área urbana do Brasil: avaliação do impacto do Programa Saúde da Família com relação ao programa tradicional de controle. **Cadernos de Saúde Pública**, v. 22, n. 5, p. 987–997, 2006.

CLEVELAND, W. S. Data Science: an Action Plan for Expanding the Technical Areas of the Field of Statistics. **International Statistical Review**, v. 69, n. 1, p. 21–26, 2001.

COELHO, T. et al. Avaliação do Grau de Completude do Cartão da Gestante de Puérperas Atendidas em um Hospital Universitário. **Revista Brasileira de Ciências da Saúde**, v. 19, n. 2, p. 117–122, 2015.

COLEONE, A. C. **Avaliação da dissipação do inseticida malation utilizado em nebulização a ultrabaixo volume no controle da dengue: avaliação ecotoxicológica e de risco ambiental**. 2014. Dissertação (Mestrado em Saúde Pública). Universidade de São Paulo, São Paulo, 2014.

COLLINS, M. H. et al. EVITA Dengue: a cluster-randomized controlled trial to Evaluate the efficacy of Wolbachia-Infected Aedes aegypti mosquitoes in reducing the incidence of Arboviral infection in Brazil. **Trials**, v. 23, n. 1, p. 185, 2022.

CONTANDRIOPOULOS, A.-P. et al. A avaliação na área da saúde: conceitos e métodos. In: HARTZ, Z.M.A.(org.). **Avaliação em saúde: dos modelos conceituais à prática na análise da implantação de programas**. Rio de Janeiro: FIOCRUZ, 2000. p. 29-47. Disponível em Scielo Books: <https://books.scielo.org/>.

CONTROLADORIA-GERAL DA UNIÃO. **5º Plano de Ação Nacional em Governo Aberto**. Disponível em: <https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp/planos-de-acao/5o-plano-de-acao-brasileiro>. Acesso em: 25 mar. 2023.

CONTROLADORIA-GERAL DA UNIÃO. **A OGP**. Disponível em: <https://www.gov.br/cgu/pt-br/governo-aberto/a-ogp>. Acesso em: 25 mar. 2023.

COSTA, C. S. et al. Programa de Controle da Esquistossomose: avaliação da implantação em três municípios da Zona da Mata de Pernambuco, Brasil. **Saúde em Debate**, v. 41, n. spe, p. 229–241, 2017.

COSTA, E. M. S.; CUNHA, R. V.; COSTA, E. A. Avaliação da implantação do Programa de Controle da Dengue em dois municípios fronteiriços do estado de Mato Grosso do Sul, Brasil, 2016*. **Epidemiologia e Serviços de Saúde**, v. 27, n. 4, p. e2017478, 2018.

DATASUS. **CNES - Estabelecimentos**. Disponível em: <https://datasus.saude.gov.br/cnes-estabelecimentos/>. Acesso em: 9 mar. 2022a.

DATASUS. **CNES - Recursos Humanos a partir de agosto de 2007**. Disponível em: <https://datasus.saude.gov.br/cnes-recursos-humanos-a-partir-de-agosto-de-2007-ocupacoes-classificadas-pela-cbo-2002>. Acesso em: 28 fev. 2022b.

DAVI, C. et al. Severe Dengue Prognosis Using Human Genome Data and Machine Learning. **IEEE Transactions on Biomedical Engineering**, v. 66, n. 10, p. 2861–2868, 2019.

DE ALMEIDA, S. P. et al. Análise da qualidade e oportunidade dos dados do SINAN no enfrentamento à dengue: reflexões e contribuições para políticas públicas de saúde. **Práticas e Cuidado: Revista de Saúde Coletiva**, v. 2, p. e12964–e12964, 2021.

DE ANDRADE, N. F. et al. Comunicação na prevenção e controle de dengue, chikungunya e Zika: um panorama analisado junto à população brasileira. **Enfermagem em Foco**, v. 12, n. 7. SUPL. 1, 2021.

DE SOUZA MELO, M. A. et al. Percepção dos profissionais de saúde sobre os fatores associados à subnotificação no Sistema Nacional de Agravos de Notificação (Sinan). **Revista de Administração em Saúde**, v. 18, n. 71, 2018.

DEHGHANI, M. H.; OMRANI, G. A.; KARRI, R. R. Solid waste—sources, toxicity, and their consequences to human health. Em: **Soft computing techniques in solid waste and wastewater management**. p. 205–213, 2021.

DEY, S. K. et al. Prediction of dengue incidents using hospitalized patients, metrological and socio-economic data in Bangladesh: A machine learning approach. **PLOS ONE**, v. 17, n. 7, p. e0270933, 2022.

DIAS, J. P. **Avaliação da efetividade do Programa de Erradicação do Aedes aegypti. Brasil, 1996-2002**. 2006. Tese (Doutorado em Saúde Pública). Universidade Federal da Bahia, Salvador, 2006.

DOMINGOS, M. D. F. et al. Atividades de controle de Aedes aegypti em municípios da Região Sul do Estado de São Paulo. **Journal of Health & Biological Sciences**, v. 1, n. 3, p. 122, 2013.

DONABEDIAN, A. The Quality of Care. **JAMA**, v. 260, n. 12, p. 1743, 1988.

DUBEUX, L. S. et al. Avaliação do Programa de Enfrentamento às Doenças Negligenciadas para o controle da esquistossomose mansônica em três municípios hiperendêmicos, Pernambuco, Brasil, 2014*. **Epidemiologia e Serviços de Saúde**, v. 28, n. 2, 2019.

DUONG, V. et al. Asymptomatic humans transmit dengue virus to mosquitoes. **Proceedings of the National Academy of Sciences**, v. 112, n. 47, p. 14688–14693, 2015.

EISEN, L. et al. Proactive Vector Control Strategies and Improved Monitoring and Evaluation Practices for Dengue Prevention. **Journal of Medical Entomology**, v. 46, n. 6, p. 1245–1255, 2009.

EL NAQA, I.; MURPHY, M. J. What Is Machine Learning? Em: **Machine Learning in Radiation Oncology**. Cham: Springer International Publishing. p. 3–11, 2015.

ELLWANGER, J. H. et al. Beyond diversity loss and climate change: Impacts of Amazon deforestation on infectious diseases and public health. **Anais da Academia Brasileira de Ciências**, v. 92, 2020.

ERLANGER, T. E.; KEISER, J.; UTZINGER, J. Effect of dengue vector control interventions on entomological parameters in developing countries: a systematic review and meta-analysis. **Medical and Veterinary Entomology**, v. 22, n. 3, p. 203–221, 2008.

ESPINAL, M. A. et al. Emerging and reemerging Aedes-transmitted arbovirus infections in the region of the Americas: implications for health policy. **American journal of public health**, v. 109, n. 3, p. 387–392, 2019.

EVANGELISTA, J. G. et al. AGENTES DE COMBATE ÀS ENDEMIAS: CONSTRUÇÃO DE IDENTIDADES PROFISSIONAIS NO CONTROLE DA DENGUE. **Trabalho, Educação e Saúde**, v. 17, n. 1, 2018.

FELICIANO, T.; CORDEIRO, B. C. Análise da qualidade dos dados das Fichas de Notificação Compulsória de Dengue e Chikungunya. **Research, Society and Development**, v. 10, n. 9, p. e40810918172–e40810918172, 2021.

FICK, S. E.; HIJMANS, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. **International Journal of Climatology**, v. 37, n. 12, p. 4302–4315, 2017.

FIGUEIREDO, L. T. M. Emergent arboviruses in Brazil. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 40, n. 2, p. 224–229, 2007.

FONSECA JÚNIOR, D. P. DA et al. Vectors of arboviruses in the state of São Paulo: 30 years of Aedes aegypti and Aedes albopictus. **Revista de Saúde Pública**, v. 53, p. 84, 2019.

FREIRE, G. N. Avaliação da qualidade dos dados notificados de dengue, macrorregião de saúde leste de Minas Gerais, 2014 a 2019. 2021.

FUNDAÇÃO SEADE. PIB. Disponível em: <https://repositorio.seade.gov.br/group/seade-pib>. Acesso em: 28 fev. 2022a.

FUNDAÇÃO SEADE. Mortalidade Infantil. Disponível em: <https://repositorio.seade.gov.br/group/seade-mortalidade>. Acesso em: 28 fev. 2022b.

FUNDAÇÃO SEADE. SEADE PAINEL. Disponível em: <https://painel.seade.gov.br/>. Acesso em: 28 fev. 2022c.

FUNDAÇÃO SEADE. SEADE Repositório. Disponível em: <https://repositorio.seade.gov.br/>. Acesso em: 19 fev. 2022d.

GABRIEL RODRIGUES DA SILVEIRA. **CARACTERIZAÇÃO DOS AGROTÓXICOS UTILIZADOS NAS AÇÕES DE CONTROLE VETORIAL E NOCIDIDADES PARA A SAÚDE DOS TRABALHADORES**. 2020. 124p. Dissertação (Mestrado em Saúde Pública e Meio Ambiente). Fundação Oswaldo Cruz, Rio de Janeiro, 2020.

GARCIA, A. L. F. et al. Self-medication and adherence to drug treatment: assessment of participants of the Universidade do Envelhecer (the University of Aging) program. **Revista Brasileira de Geriatria e Gerontologia**, v. 21, n. 6, p. 691–700, dez. 2018.

GBIF. **The Global Biodiversity Information Facility**. Disponível em: <https://www.gbif.org/what-is-gbif>. Acesso em: 25 mar. 2023.

GO FAIR BRASIL. **GO Fair Brasil**. Disponível em: <https://www.go-fair-brasil.org/>. Acesso em: 25 mar. 2023.

GODÓI, I. P. et al. Economic and epidemiological impact of dengue illness in brazil. **HTAi Cologne, Germany**, 2019.

GOMES, A. C. Vigilância entomológica. **Informe Epidemiológico do SUS**, v. 11, n. 2, p. 79–90, 2002.

GONÇALVES, I. R. et al. Avaliação do programa de uso da imunoglobulina palivizumabe no Estado de São Paulo, Brasil. **Cadernos de Saúde Pública**, v. 34, n. 7, 2018.

GOULD, E. et al. Emerging arboviruses: why today? **One health**, v. 4, p. 1–13, 2017.

GRADUS, J. L. et al. Prediction of Sex-Specific Suicide Risk Using Machine Learning and Single-Payer Health Care Registry Data From Denmark. **JAMA Psychiatry**, v. 77, n. 1, p. 25, 2020.

GUO, P. et al. Developing a dengue forecast model using machine learning: A case study in China. **PLOS Neglected Tropical Diseases**, v. 11, n. 10, p. e0005973, 16 out. 2017.

HALSTEAD, S. B. Dengue virus–mosquito interactions. **Annu. Rev. Entomol.**, v. 53, p. 273–291, 2008.

HOYOS, W.; AGUILAR, J.; TORO, M. Dengue models based on machine learning techniques: A systematic literature review. **Artificial Intelligence in Medicine**, v. 119, p. 102157, 2021.

IBGE. **Aglomerados Subnormais**. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/tipologias-do-territorio/15788-aglomerados-subnormais.html>. Acesso em: 28 jul. 2022a.

IBGE. **Regiões de influência das cidades: 2018**. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/redes-e-fluxos-geograficos/15798-regioes-de-influencia-das-cidades.html>. Acesso em: 9 mar. 2022b.

IBGE. **Cidades e Estados**. Disponível em: <https://www.ibge.gov.br/cidades-e-estados>. Acesso em: 9 mar. 2022c.

IBGE. **Estimativas da População**. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html>. Acesso em: 9 mar. 2022d.

IBGE. **PIB**. Disponível em: <https://www.ibge.gov.br/estatisticas/downloads-estatisticas.html>. Acesso em: 9 mar. 2022e.

INMON, W. H. **Building the Data Warehouse**. Fourth ed. Indianapolis: Wiley Publishing, Inc., 2005.

IPEA. **IDH 2010**. Disponível em: <https://www.ipea.gov.br/ipeageo/bases.html>. Acesso em: 9 mar. 2022.

JOSHI, A.; MILLER, C. Review of machine learning techniques for mosquito control in urban environments. **Ecological Informatics**, v. 61, p. 101241, 2021.

KATZELNICK, L. C. et al. Zika virus infection enhances future risk of severe dengue disease. **Science**, v. 369, n. 6507, p. 1123–1128, 2020.

KOSTKOVA, P. et al. Who Owns the Data? Open Data for Healthcare. **Frontiers in Public Health**, v. 4, 2016.

KRYSTOSIK, A. et al. Solid wastes provide breeding sites, burrows, and food for biological disease vectors, and urban zoonotic reservoirs: a call to action for solutions-based research. **Frontiers in public health**, v. 7, p. 405, 2020.

LEE, K. et al. Diabetic patients suffering dengue are at risk for development of dengue shock syndrome/severe dengue: emphasizing the impacts of co-existing comorbidity (ies) and glycemic control on dengue severity. **Journal of Microbiology, Immunology and Infection**, v. 53, n. 1, p. 69–78, 2020.

LEWIS, D. COVID-19 rarely spreads through surfaces. So why are we still deep cleaning. **Nature**, v. 590, n. 7844, p. 26–28, 2021.

L'HEUREUX, A. et al. Machine Learning With Big Data: Challenges and Approaches. **IEEE Access**, v. 5, p. 7776–7797, 2017.

LI, Y. et al. Urbanization increases *Aedes albopictus* larval habitats and accelerates mosquito development and survivorship. **PLoS neglected tropical diseases**, v. 8, n. 11, p. e3301, 2014.

LORICA, B. et al. **What Is a Lakehouse?** Disponível em: <https://www.databricks.com/blog/2020/01/30/what-is-a-data-lakehouse.html>. Acesso em: 31 jan. 2023.

LOWE, R. et al. Combined effects of hydrometeorological hazards and urbanisation on dengue risk in Brazil: a spatiotemporal modelling study. **The Lancet Planetary Health**, v. 5, n. 4, p. e209–e219, 2021.

MACEDO HAIR, G.; FONSECA NOBRE, F.; BRASIL, P. Characterization of clinical patterns of dengue patients using an unsupervised machine learning approach. **BMC Infectious Diseases**, v. 19, n. 1, p. 649, 2019.

MACHADO, A. A. V. Caracterização dos custos diretos e indiretos de dengue em Mato Grosso do Sul. 2019.

MACIEL, I. J.; SIQUEIRA JÚNIOR, J. B. S.; MARTELLI, C. M. T. Epidemiologia e desafios no controle do dengue. **Revista de Patologia Tropical/Journal of Tropical Pathology**, v. 37, n. 2, p. 111–130, 2008.

MADEIRA, N. G. et al. Education in primary school as a strategy to control dengue. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 35, n. 3, p. 221–226, 2002.

MAJEED, A.; LEE, S. Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey. **IEEE Access**, v. 9, p. 8512–8545, 2021.

MARENGO, J. A. et al. A seca e a crise hídrica de 2014-2015 em São Paulo. **Revista USP**, n. 106, p. 31–44, 2015.

MARQUES, C. C. DE A.; CARVALHEIRO, J. DA R. Avaliação da rede de diagnóstico laboratorial na implantação do Programa de Prevenção e Controle das Hepatites Virais no estado de São Paulo, 1997-2012*. **Epidemiologia e Serviços de Saúde**, v. 26, n. 3, p. 513–524, 2017.

MARTELLI, C. M. T. et al. Economic impact of dengue: multicenter study across four Brazilian regions. **PLoS neglected tropical diseases**, v. 9, n. 9, p. e0004042, 2015.

MEI, J.; DESROSIERS, C.; FRASNELLI, J. Machine Learning for the Diagnosis of Parkinson's Disease: A Review of Literature. **Frontiers in Aging Neuroscience**, v. 13, 2021.

MENDES, R. B. et al. Avaliação da qualidade do pré-natal a partir das recomendações do Programa de Humanização no Pré-natal e Nascimento. **Ciência & Saúde Coletiva**, v. 25, n. 3, p. 793–804, 2020.

MESSINA, J. P. et al. The current and future global distribution and population at risk of dengue. **Nature microbiology**, v. 4, n. 9, p. 1508–1515, 2019.

MHASAWADE, V.; ZHAO, Y.; CHUNARA, R. Machine learning and algorithmic fairness in public and population health. **Nature Machine Intelligence**, v. 3, n. 8, p. 659–666, 2021.

MILOSLAVSKAYA, N.; TOLSTOY, A. Big Data, Fast Data and Data Lake Concepts. **Procedia Computer Science**, v. 88, p. 300–305, 2016.

MINISTÉRIO DA SAÚDE. **Instruções para Pessoal de Combate ao Vetor**. Disponível em: https://bvsmms.saude.gov.br/bvs/publicacoes/funasa/man_dengue.pdf. Acesso em: 28 fev. 2022a.

MINISTÉRIO DA SAÚDE. **Programa Nacional de Controle da Dengue (PNCD)**. Disponível em: https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/svsa/dengue/pncd_2002.pdf/view. Acesso em: 28 fev. 2022b.

MINISTÉRIO DA SAÚDE. **Diretrizes Nacionais para a Prevenção e Controle de Epidemias de Dengue**. Disponível em: https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/svsa/dengue/diretrizes_nacionais_prevencao_controle_dengue.pdf/view. Acesso em: 28 fev. 2022c.

MINISTÉRIO DA SAÚDE. **Doenças Infecciosas e Parasitárias: Guia de Bolso**, 2010. Disponível em: https://bvsmis.saude.gov.br/bvs/publicacoes/doencas_infecciosas_parasitaria_guia_bolso.pdf. Acesso em: 28 fev. 2022d

MINISTÉRIO DA SAÚDE. **Política Nacional de Atenção Básica**. Disponível em: https://bvsmis.saude.gov.br/bvs/saudelegis/gm/2017/prt2436_22_09_2017.html. Acesso em: 28 fev. 2022e.

MINISTÉRIO DA SAÚDE. **Sistema de Informações sobre Orçamentos Públicos em Saúde**. Disponível em: <https://www.gov.br/saude/pt-br/acao-saude/informacao/siops>. Acesso em: 28 fev. 2022f.

MORIN, C. W.; COMRIE, A. C.; ERNST, K. Climate and dengue transmission: evidence and implications. **Environmental health perspectives**, v. 121, n. 11–12, p. 1264–1272, 2013.

MULLIGAN, K. et al. Is dengue a disease of poverty? A systematic review. **Pathogens and global health**, v. 109, n. 1, p. 10–18, 2015.

MYERS, S. S.; PIVOR, J. I.; SARAIVA, A. M. The São Paulo Declaration on Planetary Health. **The Lancet**, v. 398, n. 10308, p. 1299, 2021.

NAMBIAR, A.; MUNDRA, D. An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management. **Big Data and Cognitive Computing**, v. 6, n. 4, p. 132, 2022.

NEGASH, S. Business Intelligence. **Communications of the Association for Information Systems**, v. 13, 2004.

NEIDERUD, C.-J. How urbanization affects the epidemiology of emerging infectious diseases. **Infection ecology & epidemiology**, v. 5, n. 1, p. 27060, 2015.

NICHOLS, G.; LAKE, I.; HEAVISIDE, C. Climate change and water-related infectious diseases. **Atmosphere**, v. 9, n. 10, p. 385, 2018.

NOVAES, H. M. D. Avaliação de programas, serviços e tecnologias em saúde. **Revista de Saúde Pública**, v. 34, n. 5, p. 547–549, 2000.

OHDSI. **Standardized Data: The OMOP Common Data Model**. Disponível em: <https://www.ohdsi.org/data-standardization/>. Acesso em: 28 fev. 2022.

PAN AMERICAN HEALTH ORGANIZATION. **Dengue in the Americas reaches highest number of cases recorded**. Disponível em: https://www3.paho.org/hq/index.php?option=com_content&view=article&id=15593:dengue-in-the-americas-reaches-highest-number-of-cases-recorded&Itemid=0&lang=en#gsc.tab=0. Acesso em: 28 fev. 2022a.

PAN AMERICAN HEALTH ORGANIZATION. **Epidemiological Update: Dengue. 11**. Disponível em: https://www3.paho.org/hq/index.php?option=com_docman&view=download&alias=50963-11-november-2019-dengue-epidemiological-update-1&category_slug=dengue-2217&Itemid=270&lang=en. Acesso em: 28 fev. 2022b.

PAN AMERICAN HEALTH ORGANIZATION. **Epidemiological Update: Arbovirus**. Disponível em: <https://www3.paho.org/data/index.php/en/mnu->

topics/indicadores-dengue-en/annual-arbovirus-bulletin-2020.html. Acesso em: 28 fev. 2022c.

PATEL, J. Bridging Data Silos Using Big Data Integration. **International Journal of Database Management Systems**, v. 11, n. 3, p. 01–06, 2019.

PESSANHA, J. E. M. et al. Avaliação do Plano Nacional de Controle da Dengue. **Cadernos de Saúde Pública**, v. 25, n. 7, p. 1637–1641, 2009.

PESSOA, J. P. M. et al. Controle da dengue: os consensos produzidos por Agentes de Combate às Endemias e Agentes Comunitários de Saúde sobre as ações integradas. **Ciência & Saúde Coletiva**, v. 21, n. 8, p. 2329–2338, 2016.

PORRUA, P. et al. Avaliação da Gestão do Programa Nacional de Alimentação Escolar sob a ótica do fomento da agricultura familiar. **Cadernos Saúde Coletiva**, v. 28, n. 3, p. 426–437, 2020.

QUARESMA, M. A. **Avaliação da implantação do componente controle vetorial do Programa de Controle da Dengue em Porto Seguro- BA/Brasil**. Rio de Janeiro: Fundação Oswaldo Cruz, 2017.

RANK, A. T.; BERBERI, M. A. L. Big data e direitos fundamentais sob o enfoque da Lei Geral de Proteção de Dados (LGPD). **International Journal of Digital Law**, v. 3, n. 2, p. 9–26, 2022.

RAVAUT, M. et al. Predicting adverse outcomes due to diabetes complications with machine learning using administrative health data. **npj Digital Medicine**, v. 4, n. 1, p. 24, 2021.

RAVI, D. et al. Deep Learning for Health Informatics. **IEEE Journal of Biomedical and Health Informatics**, v. 21, n. 1, p. 4–21, 2017.

RAWLINSON, S.; CIRIC, L.; CLOUTMAN-GREEN, E. COVID-19 pandemic—let’s not forget surfaces. **Journal of hospital infection**, v. 105, n. 4, p. 790–791, 2020.

ROHANI, A. et al. The effect of extrinsic incubation temperature on development of dengue serotype 2 and 4 viruses in *Aedes aegypti* (L.). **Southeast Asian Journal of Tropical Medicine and Public Health**, v. 40, n. 5, p. 942, 2009.

ROLIM, G. S. et al. **Atualização da classificação climática de Köppen e de Thornthwaite para o Estado de São Paulo com o uso de sistema de informação geográfica**. 14 CONGRESSO BRASILEIRO DE AGROMETEOROLOGIA. Anais. Campinas: 2005.

SAGIROGLU, S.; SINANC, D. **Big data: A review**. 2013 International Conference on Collaboration Technologies and Systems (CTS). Anais. IEEE, 2013.

SALIM, N. A. M. et al. Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. **Scientific Reports**, v. 11, n. 1, p. 939, 2021.

SANTOS, D. R. et al. Avaliação da eficácia do Programa Rede Mãe Paranaense. **Saúde em Debate**, v. 44, n. 124, p. 70–85, 2020.

SECRETARIA DE ESTADO DA SAÚDE DE SÃO PAULO. **SUCEN - Superintendência de Controle de Endemias**. Disponível em: <https://www.saude.sp.gov.br/sucen-superintendencia-de-controle-de-endemias/>. Acesso em: 28 fev. 2022a.

SECRETARIA DE SAÚDE DE SÃO PAULO. **Diretrizes para a prevenção e controle das arboviroses urbanas no estado de São Paulo**. Disponível em: <https://cvs.saude.sp.gov.br/up/Diretrizes%20controle%20arboviroses%20ESP%20-%202017.pdf>. Acesso em: 28 fev. 2022b.

SECRETARIA DE VIGILÂNCIA EM SAÚDE. **Mortalidade infantil no Brasil**. Brasília. Disponível em: <https://saude.rs.gov.br/upload/arquivos/202106/11173526-boletim-epidemiologico-mortalidade-materna-e-mortalidade-infantil-2021.pdf>. Acesso em: 28 fev. 2023.

SHCHERBAKOV, M. et al. Lean Data Science Research Life Cycle: A Concept for Data Analysis Software Development. p. 708–716.

SILVA, A. A. et al. Avaliação do Sistema de Vigilância do Programa Nacional de Imunizações - Módulo Registro do Vacinado, Brasil, 2017. **Epidemiologia e Serviços de Saúde**, v. 30, n. 1, 2021.

SILVA, N. P. **A utilização dos programas TABWIN e TABNET como ferramentas de apoio a disseminação das informações em saúde**. Rio de Janeiro: Fundação Oswaldo Cruz, 2009.

SINAN. **Doenças e Agravos de Notificação - 2007 em diante (SINAN)**. Disponível em: <https://datasus.saude.gov.br/aceso-a-informacao/doencas-e-agravos-de-notificacao-de-2007-em-diante-sinan/>. Acesso em: 28 fev. 2023.

SISTEMA NACIONAL DE INFORMAÇÕES SOBRE SANEAMENTO (SNIS). **Painel de Saneamento**. Disponível em: <https://www.gov.br/mdr/pt-br/assuntos/saneamento/snis>. Acesso em: 28 fev. 2023.

SOUZA, C. M. et al. Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine. **Remote Sensing**, v. 12, n. 17, p. 2735, 2020.

SOUZA, M. F. et al. Avaliação da implantação do Programa de Controle da Hanseníase em Camaragibe, Pernambuco*. **Epidemiologia e Serviços de Saúde**, v. 26, n. 4, p. 817–834, 2017.

SOUZA, K. R. et al. Saberes e práticas sobre controle do *Aedes aegypti* por diferentes sujeitos sociais na cidade de Salvador, Bahia, Brasil. **Cadernos de Saúde Pública**, v. 34, n. 5, p. e00078017, 2018.

SYAKUR, M. A. et al. Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. **IOP Conference Series: Materials Science and Engineering**, v. 336, p. 012017, 2018.

SYLVESTRE, E. et al. Data-driven methods for dengue prediction and surveillance using real-world and Big Data: A systematic review. **PLOS Neglected Tropical Diseases**, v. 16, n. 1, p. e0010056, 2022.

TANTOWIJOYO, W. et al. *Aedes aegypti* abundance and insecticide resistance profiles in the Applying Wolbachia to Eliminate Dengue trial. **PLOS Neglected Tropical Diseases**, v. 16, n. 4, p. e0010284, 2022.

TEIXEIRA, M. G.; BARRETO, M. L.; GUERRA, Z. Epidemiologia e medidas de prevenção do dengue. **Informe epidemiológico do SUS**, v. 8, n. 4, p. 5–33, 1999.

TEIXEIRA, M. G. et al. Dengue: twenty-five years since reemergence in Brazil. **Cadernos de saúde pública**, v. 25, p. S7–S18, 2009.

TREVEIL, M. et al. **Introducing MLOps**. O'Reilly Media, 2020.

UCAR. **Network Common Data Form (NetCDF)**. Disponível em: <https://www.unidata.ucar.edu/software/netcdf/>. Acesso em: 31 jan. 2023.

UNITED NATIONS. **World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420)**. Disponível em: <https://www.un.org/en/desa/2018-revision-world-urbanization-prospects#:~:text=The%20urban%20population%20of%20the,and%20Africa%20with%2013%25%20each..> Acesso em: 28 fev. 2022.

VAICHULONIS, C. G. et al. Evaluation of prenatal care according to indicators for the Prenatal and Birth Humanization Program. **Revista Brasileira de Saúde Materno Infantil**, v. 21, n. 2, p. 441–450, 2021.

VALLE, D.; BELINATO, T. A.; MARTINS, A. J. Controle químico de *Aedes aegypti*, resistência a inseticidas e alternativas. **Dengue: teorias e práticas Rio de Janeiro: Fiocruz**, p. 93–126, 2015.

VAN DER AALST, W. M. P. Data Scientist: The Engineer of the Future. Em: **Enterprise Interoperability VI**. Cham: Springer International Publishing, 2014. p. 13–26.

VARGAS, L. D. L. et al. Resistência das populações de *Aedes (Stegomyia) aegypti* (Linnaeus, 1762)(Insecta, Diptera, Culicidae) aos inseticidas utilizados para o controle: estado da arte do conhecimento. **Rev. Ciênc. Méd. Biol.(Impr.)**, p. 98–116, 2022.

VEIGA, V. S. de O. et al. **Relatório de atividades Rede GO FAIR Brasil Saúde 2018/2021**. Rio de Janeiro: Fiocruz/ICICT, 2021. 29 p.

VON SPERLING, O.; LADEIRA, M. **Mining Twitter Data for Signs of Depression in Brazil**. In: Anais do VII Symposium on Knowledge Discovery, Mining and Learning (KDMiLe), 2019, Fortaleza. **Anais**. Porto Alegre.Sociedade Brasileira de Computação – SBC, 2019. p. 25-32.

WEN, T.-H.; HSU, C.-S.; HU, M.-C. Evaluating neighborhood structures for modeling intercity diffusion of large-scale dengue epidemics. **International journal of health geographics**, v. 17, n. 1, p. 1–15, 2018.

WHITMEE, S. et al. Safeguarding human health in the Anthropocene epoch: report of The Rockefeller Foundation–Lancet Commission on planetary health. **The Lancet**, v. 386, n. 10007, p. 1973–2028, 2015.

WILKINSON, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, v. 3, n. 1, p. 160018, 15 mar. 2016.

WORLD ECONOMIC FORUM. **Jobs of Tomorrow Mapping Opportunity in the New Economy**. Genebra: World Economic Forum, 2020.

WORLD HEALTH ORGANIZATION. **Cleaning and disinfection of environmental surfaces in the context of COVID-19**. 2020.

WORLD HEALTH ORGANIZATION. **Dengue and severe dengue**. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>. Acesso em: 28 fev. 2022a.

WORLD HEALTH ORGANIZATION. **Japanese encephalitis**. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/japanese-encephalitis>. Acesso em: 28 fev. 2022b.

ZARA, A. L. DE S. A. et al. Estratégias de controle do *Aedes aegypti*: uma revisão. **Epidemiologia e Serviços de Saúde**, v. 25, n. 2, p. 1–2, 2016.

ZOABI, Y.; DERI-ROZOV, S.; SHOMRON, N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. **npj Digital Medicine**, v. 4, n. 1, p. 3, 2021.

Apêndice A - Artefatos desenvolvidos

Os scripts Python desenvolvidos para esta pesquisa estão versionados em <https://github.com/professorxavier/denguetese>.