

Investigando Fragilidades na Coleta de Dados em Sistemas de Informação em Saúde e suas Implicações para Modelagens e a Tomada de Decisão no SUS

Flávia Rosane de Mendonça Luis¹, Thiago Meirelles Ventura¹

¹Instituto de Computação – Universidade Federal de Mato Grosso (UFMT)
Av. Fernando Corrêa da Costa, no 2367, Bairro Boa Esperança, 78.060-900, Cuiabá
MT – Brasil

flavia.luis@sou.ufmt.br, thiago@ic.ufmt.br

***Abstract.** This study investigates weaknesses in the collection and storage of data in health information systems and their implications for data modeling and decision-making. The findings indicate that human, organizational, and technological aspects may lead to inconsistencies and inaccuracies in the data. As next steps, questionnaires will be applied to health professionals to identify perceived failures in data collection and storage, and simulations of health data modeling will be conducted to evaluate the impact of these failures on analyses and decision-making.*

***Resumo.** Este estudo investiga fragilidades na coleta e no armazenamento de dados em sistemas de informação em saúde e suas possíveis implicações para modelagens de dados e tomada de decisão clínica e gerencial. Os resultados indicam que aspectos humanos, organizacionais e tecnológicos podem gerar inconsistências e imprecisões nos dados. Como continuidade da pesquisa, serão aplicados questionários com profissionais de saúde para identificar falhas percebidas na coleta e armazenamento de dados, além da realização de simulações de modelagens incorporando essas falhas para avaliar seus impactos nas análises e decisões.*

1. Introdução

O Sistema Único de Saúde (SUS) representa um marco na garantia do direito universal à saúde no Brasil, sendo estruturado a partir dos princípios de universalidade, integralidade e equidade. Sua organização ocorre por meio de níveis progressivos de atenção, que compreendem a Atenção Primária, Secundária e Terciária, articulados em uma rede de serviços voltada à promoção, prevenção e recuperação da saúde da população [Brasil 1988]. Nesse contexto, a Atenção Primária à Saúde desempenha papel central como porta de entrada do sistema e responsável pela coordenação do cuidado, especialmente por meio da Estratégia Saúde da Família, que fortalece a relação entre os profissionais de saúde e as comunidades atendidas [Brasil 2006].

A gestão desse sistema depende fortemente da produção e do uso de informações confiáveis. Com o avanço da digitalização dos processos administrativos e assistenciais, os Sistemas de Informação em Saúde passaram a desempenhar papel estratégico no registro, organização e disponibilização de dados relacionados aos atendimentos, aos procedimentos e à vigilância epidemiológica. A utilização dessas informações possibilita

a construção de históricos clínicos mais completos, além de apoiar atividades de planejamento, monitoramento e avaliação das ações em saúde [De Oliveira 2024]. Contudo, a existência de múltiplos sistemas desenvolvidos de forma independente no contexto do SUS levanta a hipótese de que a fragmentação tecnológica possa dificultar a integração das informações e comprometer o uso mais amplo dos dados disponíveis [Coelho Neto e Chioro 2021].

Além dos desafios relacionados à estrutura dos sistemas, também se supõe que fragilidades associadas ao processo de coleta e registro das informações possam impactar a qualidade dos dados produzidos. Aspectos como limitações de infraestrutura, ausência de padronização nos procedimentos de registro, lacunas na capacitação profissional e fatores comportamentais relacionados à rotina de trabalho dos profissionais de saúde são frequentemente apontados na literatura como possíveis obstáculos à produção de informações consistentes e completas [Amare 2025], [Salangwa 2025]. Nesse cenário, existe ainda a possibilidade de que determinadas informações sejam registradas de forma incompleta ou estimada, especialmente em contextos de alta demanda assistencial ou quando o preenchimento de sistemas ocorre predominantemente para fins administrativos [Tan 2023].

Considerando a crescente utilização de dados para subsidiar análises epidemiológicas, modelagens computacionais e processos de tomada de decisão em saúde, compreender essas possíveis fragilidades torna-se fundamental. Assim, este trabalho tem como objetivo investigar fragilidades na coleta de dados em sistemas de informação em saúde e discutir suas possíveis implicações para a modelagem de dados e para a tomada de decisão no contexto do SUS.

2. Materiais e Métodos

Para fundamentar a investigação acerca do impacto da qualidade dos dados nos diferentes níveis de atenção à saúde, foi conduzida uma revisão sistemática da literatura estruturada, seguindo o protocolo descrito em Kitchenham [Kitchenham 2024].

Na etapa de planejamento foram definidas as questões de pesquisa que orientaram o levantamento bibliográfico, bem como os critérios de inclusão e exclusão dos estudos. Os critérios de inclusão consideraram trabalhos que abordassem aspectos relacionados à coleta de dados em saúde, à qualidade dos dados ou às implicações desses fatores para a gestão e a tomada de decisão em saúde. Foram excluídos estudos que não apresentavam relação direta com o processo de coleta de dados ou que não tratavam do contexto de sistemas de informação em saúde.

O levantamento dos trabalhos foi realizado a partir da definição da estratégia PICOC (População, Intervenção, Comparação, Outcomes/Resultados e Contexto), por meio de buscas em bases de dados científicas, utilizando combinações de termos relacionados a sistemas de informação em saúde, coleta de dados, qualidade da informação e tomada de decisão. As estratégias de busca foram definidas de modo a ampliar a cobertura da literatura relevante, contemplando diferentes perspectivas sobre o tema. Após a obtenção dos resultados iniciais, os estudos identificados passaram por um processo de triagem, que incluiu a leitura dos títulos e resumos para verificação de aderência aos objetivos da pesquisa. O detalhamento das etapas está ilustrado na Figura 1.

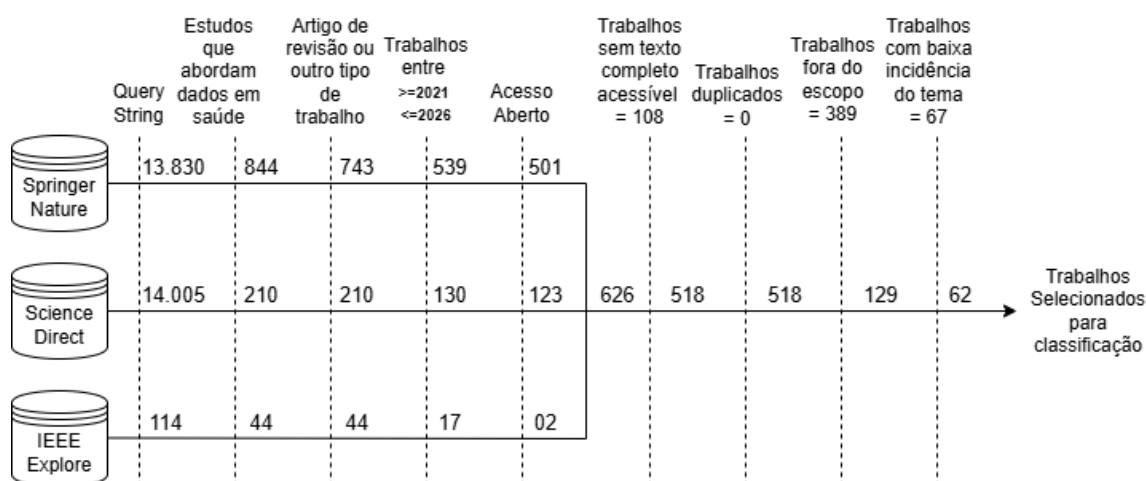


Figura 1 – Seleção dos trabalhos relacionados à pesquisa

Em seguida, foi realizada a leitura completa dos 62 trabalhos selecionados, permitindo a extração das informações relevantes para a análise. Os estudos incluídos foram então examinados com o objetivo de identificar as principais fragilidades associadas à coleta de dados em saúde, bem como as implicações apontadas na literatura para a utilização dessas informações em processos analíticos e decisórios.

3. Resultados e Discussões

A análise dos estudos selecionados na revisão sistemática evidenciou que as fragilidades na coleta de dados em sistemas de informação em saúde estão associadas a diferentes dimensões organizacionais, tecnológicas e comportamentais. Um dos aspectos recorrentes na literatura refere-se à fragmentação dos Sistemas de Informação em Saúde, caracterizada pela existência de múltiplos sistemas desenvolvidos para finalidades específicas e com níveis limitados de interoperabilidade [Eisinger-Mathason 2025]. Essa estrutura dificulta a integração dos dados e a construção de uma visão consolidada do histórico de saúde dos usuários, comprometendo o acompanhamento longitudinal do cuidado e o monitoramento de indicadores assistenciais [Coelho Neto e Chioro 2021], [Byrne e Heywood 2023].

Outra fragilidade está relacionada ao próprio processo de registro das informações nos serviços de saúde. Há limitações de infraestrutura tecnológica, rotatividade de profissionais e ausência de treinamento adequado influenciam diretamente a qualidade dos dados produzidos. Nessas condições, o registro de informações pode ocorrer de forma incompleta ou inconsistente, reduzindo o potencial de utilização dos dados para atividades de planejamento e gestão [Daneshkohan et al. 2022], [Chanyalewet al. 2023]. Além disso, fatores comportamentais também desempenham papel relevante nesse processo, uma vez que a sobrecarga de trabalho e a baixa priorização do registro de dados podem levar ao preenchimento tardio ou impreciso das informações [Amare 2025].

Em determinados contextos, lacunas de informação são preenchidas por estimativas ou valores aproximados, especialmente quando há exigências administrativas relacionadas ao envio de dados para sistemas nacionais. Esse tipo de prática pode comprometer a confiabilidade das bases de dados utilizadas para análises epidemiológicas e planejamento de políticas públicas [Tan 2023]. Situações semelhantes foram observadas em estudos que investigaram registros eletrônicos de saúde, nos quais inconsistências e

valores incorretos foram identificados em variáveis clínicas relevantes, como altura e peso dos pacientes, demonstrando os desafios associados à garantia da qualidade dos dados em larga escala [Guide et al. 2024].

Outro conjunto de resultados refere-se às implicações dessas fragilidades para o uso secundário dos dados em atividades analíticas e computacionais. Problemas como dados ausentes, inconsistentes ou excessivamente redundantes podem comprometer o desempenho de métodos de mineração de dados e de modelos de aprendizado de máquina aplicados ao domínio da saúde. Estudos indicam que a presença de dados incompletos em registros eletrônicos pode afetar significativamente a validade das análises e gerar vieses nos resultados obtidos [Getzen et al. 2023]. De forma semelhante, a presença de registros extensos e pouco estruturados pode prejudicar o desempenho de modelos de processamento de linguagem natural voltados à predição clínica [Liu et al. 2022].

No cenário brasileiro, Luis et al. (2025) identifica uma lacuna crítica na formação dos Agentes Comunitários de Saúde (ACS), profissionais que atuam diretamente na coleta e manejo de dados sensíveis, incluindo informações demográficas, condições de saúde, histórico familiar e aspectos socioeconômicos, sem receberem capacitação adequada sobre a Lei Geral de Proteção de Dados (LGPD), evidenciando a necessidade de revisão curricular e implementação de programas permanentes de formação.

Por fim, alguns estudos destacam que a coleta e organização de dados em saúde envolve um volume significativo de atividades frequentemente invisíveis no fluxo de trabalho dos profissionais, caracterizadas como trabalho oculto de informação. Essas atividades incluem validação, correção e reorganização de registros para torná-los utilizáveis em análises e relatórios institucionais, o que evidencia a complexidade do processo de produção de dados em ambientes de saúde [Mcvey et al. 2023]. Em conjunto, esses resultados indicam que as fragilidades na coleta de dados não se limitam a aspectos tecnológicos, mas envolvem também fatores organizacionais, operacionais e humanos que influenciam diretamente a qualidade das informações disponíveis para a gestão e para a tomada de decisão no sistema de saúde.

4. Considerações Finais

Este trabalho teve como objetivo investigar fragilidades na coleta de dados em sistemas de informação em saúde e discutir suas possíveis implicações para a modelagem de dados e para os processos de tomada de decisão no contexto do Sistema Único de Saúde. Para isso, foi conduzida uma revisão sistemática da literatura, buscando identificar evidências científicas que permitissem compreender os principais desafios associados à produção e ao uso de informações em saúde.

Os resultados obtidos indicam que as fragilidades na coleta de dados estão associadas a múltiplos fatores. Entre eles destacam-se a fragmentação dos sistemas de informação, as limitações de interoperabilidade entre plataformas, problemas estruturais relacionados à infraestrutura e à capacitação dos profissionais, além de aspectos comportamentais ligados à rotina de trabalho nos serviços de saúde. Há também evidências que práticas como o registro incompleto ou estimado de informações podem comprometer a qualidade das bases de dados utilizadas em análises epidemiológicas, na gestão dos serviços e no desenvolvimento de modelos computacionais aplicados à área da saúde.

Essas limitações tornam-se particularmente relevantes diante da crescente utilização de dados para apoiar processos analíticos, modelagens computacionais e atividades de planejamento em saúde pública. Dados inconsistentes, incompletos ou pouco padronizados podem introduzir vieses nas análises e reduzir a confiabilidade das informações utilizadas para subsidiar decisões clínicas e gerenciais. Dessa forma, compreender as fragilidades existentes no processo de coleta de dados constitui um passo importante para o desenvolvimento de estratégias que promovam a melhoria da qualidade das informações em sistemas de saúde.

Como continuidade desta pesquisa, duas etapas adicionais estão previstas. A primeira consiste na aplicação de um questionário junto a agentes e profissionais da saúde, com o objetivo de compreender a percepção desses atores em relação às possíveis falhas presentes nos processos de coleta e armazenamento de dados em sistemas de informação em saúde. A realização dessa etapa já foi submetida para apreciação do Comitê de Ética em Pesquisa (CEP). A segunda etapa envolve a realização de simulações de modelagem de dados em saúde nas quais serão inseridas, de forma controlada, as falhas identificadas na literatura e relatadas pelos profissionais. Essas simulações permitirão mensurar o impacto dessas fragilidades sobre os resultados das modelagens e, conseqüentemente, sobre os processos de tomada de decisão clínica e gerencial, contribuindo para uma compreensão mais aprofundada dos riscos associados à baixa qualidade dos dados em saúde.

Agradecimentos

O presente estudo foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado de Mato Grosso (FAPEMAT), processo FAPEMAT-PRO.0000138/2025.

Referências

- Amare, G., Shentema, M. G., Teshome, F., Tadesse, Y., Enyew, D. B., Tsegaw, D., Muleta, M. (2025). Behavioral factors affecting data quality and information utilization in health facilities in ethiopia, 2023: an exploratory qualitative study. *Archives of Public Health*, v. 83, p. 230.
- Brasil (1988). Constituição da República Federativa do Brasil de 1988”. Brasília, DF: Presidência da República.
- Brasil (2006). 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.
- Byrne, E;; Heywood, A; (2023). Use of routine health information systems data in developing and monitoring district and facility health plans: a scoping review. *BMC Health Services Research*v. 23, n. 1, p. 1049.
- Chanyalew, M. A., Yitayal, M., Atnafu, A., & Tilahun, B. (2023). Assessment of data demand for informed-decisions among health facility and department heads in public health facilities of Amhara Region, northwest Ethiopia. *Health research policy and systems*, v. 21, n. 1, p. 62, 2023.

- Coelho Neto, Giliate,, C.; Chioro, A. (2021). Afinal, quantos Sistemas de Informação em Saúde de base nacional existem no Brasil?. *Cadernos de Saúde Pública* , v. 37, p. e00182119.
- Daneshkohan, A., Alimoradi, M., Ahmadi, M., & Alipour, J. (2022). Data quality and data use in primary health care: A case study from Iran. *Informatics in Medicine Unlocked*, v. 28, p. 100855.
- De Oliveira, L. S., Silva, G. G., Magalhães, E. V., Moura, O. F., Schettert, I. R., Simon, O. A. A., Paula, B. C., Lopes, A. L. B., Cunha H. C. (2024). Gestão eficiente da atenção básica no sus: rumo a um sistema de saúde mais equitativo e inclusivo. *Licença creative commons*, p. 93.
- Eisinger-Mathason, T. K., Leshin, J., Lahoti, V., Fridsma, D. B., Mucaj, V., & Kho, A. N. (2025). Data linkage multiplies research insights across diverse healthcare sectors. *Communications Medicine* v. 5, n. 1, p. 58.
- Getzen, E., Ungar, L., Mowery, D., Jiang, X., Long, Q. (2023). Mining for equitable health: Assessing the impact of missing data in electronic health records. *Journal of biomedical informatics*. v. 139, p. 104269.
- Guide, A., Sulieman, L., Garbett, S., Cronin, R. M., Spotnitz, M., Natarajan, K., Chen, Q. (2024). Identifying erroneous height and weight values from adult electronic health records in the All of Us research program. *Journal of biomedical informatics* v. 155, p. 104660.
- Kitchenham, B. Procedures for performing systematic reviews. Keele, UK, Keele Univ., v. 33, 08 2004.
- Liu, J., Capurro, D., Nguyen, A., & Verspoor, K. (2022). Note Bloat, impacts deep learning-based NLP models for clinical prediction tasks. v. 133, p. 104149.
- Luis, F. R. M., Araujo, N. V., Neto, C. D., & Ventura, T. M. (2025). Proposta de Aplicação da LGPD na Coleta e Manipulação de Dados Sensíveis na Atenção Primária do SUS para Levantamento de Requisitos dos Softwares. *Escola Regional de Sistemas de Informação de Mato Grosso*. pp. 159-164.
- Mcvey, L., Alvarado, N., Greenhalgh, J., Elshehaly, M., Gale, C. P., Lake, J., ... & Randell, R. (2021). Hidden labour: the skilful work of clinical audit data collection and its implications for secondary use of data via integrated health IT. *BMC Health Services Research*, v. 21, n. 1, p. 702.
- Salangwa, C., Munthali, R., Mfunu, L., & Nyirenda, VK (2025). Utilisation of data and factors affecting the performance of the Health Management Information System (HMIS). The case of six Synods of Livingstonia health facilities in Malawi. *Archives of Public Health*, v. 83, n. 1, p. 1-9.
- Tan, AL, Getzen, EJ, Hutch, MR, Strasser, ZH, Gutiérrez-Sacristán, A., Le, TT, Holmes, JH (2023). Informative missingness: What can we learn from patterns in missing laboratory data in the electronic health record?. *Journal of biomedical informatics*v. 139, p. 104306.