Journal of Health Informatics

J Health Inform. 2026;18:e1845
DOI: https://doi.org/10.59681/2175-4411.v18.iEspecial.2026.1845

ARTIGO COMPLETO
CBIS 2026

Panorama do uso secundário de dados em registros eletrônicos de saúde: uma revisão de escopo

Overview of the secondary use of data in electronic health records: a scoping review

Darcio Leite Mario[1] ✉, Rodrigo Paula Campos[2], Aluizio Nunes da Silva Junior[3], Marco de Oliveira da Silva[3], José Viterbo Filho[3]

ORCID: 0009-0000-3431-0697 · 0000-0001-5419-964X · 0009-0003-8803-2153 · 0009-0009-9236-5587 · 0000-0002-0339-6624

[1] HU Brasil, Hospital Universitário Antônio Pedro, Niterói (RJ), Brasil.
[2] Hospital Universitário Antônio Pedro, Universidade Federal Fluminense, Niterói (RJ), Brasil.
[3] Programa de Pós-Graduação em Computação, Universidade Federal Fluminense, Niterói (RJ), Brasil.

FICHA DO ARTIGO

TIPO
Artigo completo · CBIS 2026

IDENTIFICADOR
e1845

DOI
10.59681/2175-4411.v18.iEspecial.2026.1845

RECEBIDO
17/05/2026

APROVADO
20/07/2026

PROCESSO DE AVALIAÇÃO
Double Blind Review

EDITOR RESPONSÁVEL
Juliano de Souza Gaspar

ACESSO
Aberto, CC BY 4.0

AUTOR CORRESPONDENTE
Darcio Leite Mario
darcio.mario@ebserh.gov.br

LICENÇA

Este é um artigo publicado em acesso aberto sob a licença Creative Commons CC BY 4.0, que permite uso, distribuição e reprodução em qualquer meio, desde que o trabalho original seja corretamente citado.

CC BY 4.0

RESUMO

Objetivo: Mapear como a literatura descreve os usos secundários de dados rotineiramente coletados em registros eletrônicos de saúde (RES), com ênfase nas finalidades de reutilização e no papel da interoperabilidade, da estruturação da informação e da qualidade do dado. **Método:** Revisão de escopo orientada pelo *Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews* (PRISMA-ScR), com buscas em Scopus, PubMed/MEDLINE e Biblioteca Virtual em Saúde, triagem no Parsif.al e síntese narrativa. **Resultados:** 45 registros foram importados, 24 duplicatas removidas e 9 estudos integraram a síntese principal. Os usos identificados concentraram-se em cinco eixos: qualificação dos dados, integração multicêntrica, vigilância e evidências de mundo real, pesquisa clínica e pesquisa em serviços de saúde. **Conclusão:** A reutilização de dados em saúde depende não apenas da disponibilidade de bases digitais, mas também de interoperabilidade, padronização, rastreabilidade, transparência metodológica e governança do processamento.

Descritores: Dados de Saúde Coletados Rotineiramente; Interoperabilidade da Informação em Saúde; Registros Eletrônicos de Saúde.

ABSTRACT

Objective: To map how the literature describes the secondary uses of routinely collected data from electronic health records (EHR), with emphasis on reuse purposes and the role of interoperability, information structuring and data quality. **Method:** Scoping review guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR), with searches in Scopus, PubMed/MEDLINE and the Virtual Health Library, screening in Parsif.al and narrative synthesis. **Results:** 45 records were imported, 24 duplicates were removed, and 9 studies composed the main synthesis. The identified uses were concentrated in five axes: data qualification, multicenter integration, surveillance and real-world evidence, clinical research, and health services research. **Conclusion:** Secondary data use in health depends not only on the availability of digital databases, but also on interoperability, standardization, traceability, methodological transparency and governance of processing.

Keywords: Routinely Collected Health Data; Health Information Interoperability; Electronic Health Records.

CBIS 2026 XXI CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE

Artigo apresentado no XXI Congresso Brasileiro de Informática em Saúde (CBIS 2026), realizado de 23 a 25 de setembro de 2026, em Brasília (DF), e publicado na seção CBIS 2026 do Journal of Health Informatics.

COMO CITAR ESTE ARTIGO: Mario DL, Campos RP, da Silva Junior AN, da Silva MO, Viterbo Filho J. Panorama do uso secundário de dados em registros eletrônicos de saúde: uma revisão de escopo. J Health Inform 2026;18:e1845. doi: https://doi.org/10.59681/2175-4411.v18.iEspecial.2026.1845

1 INTRODUÇÃO

A crescente digitalização dos serviços de saúde ampliou a produção, o armazenamento e a circulação de dados clínicos e administrativos em diferentes contextos assistenciais. Nesse cenário, os registros eletrônicos de saúde (RES) passaram a ter importância não apenas para o cuidado direto, mas também para pesquisa clínica, vigilância, avaliação de serviços e outras finalidades analíticas. Desde o início da década passada, a literatura já apontava que a heterogeneidade entre sistemas, esquemas de dados e terminologias constitui uma barreira importante ao reaproveitamento de dados rotineiramente coletados para pesquisa, especialmente em contextos multicêntricos e de elegibilidade para estudos clínicos [1].

De modo geral, o uso secundário de dados em saúde pode ser entendido como a reutilização de dados originalmente produzidos no contexto assistencial para finalidades distintas do cuidado direto e imediato ao paciente. Esse reaproveitamento pode apoiar, entre outros objetivos, a construção de registros clínicos, a identificação de casos, a pesquisa epidemiológica, a pesquisa clínica, a avaliação de serviços e o monitoramento de indicadores. No entanto, a literatura também mostra que esse potencial depende de condições técnicas e informacionais específicas, como completude, consistência, ausência de duplicidades, qualidade semântica e adequação dos dados ao propósito de uso [2].

Nesse processo, a interoperabilidade assume papel central. Mais do que permitir intercâmbio técnico entre sistemas, ela precisa sustentar interpretação consistente dos dados em contextos distintos de reutilização. Além disso, o uso secundário de dados não depende apenas da existência de bases digitais, mas também da compreensão de como esses dados foram produzidos, transformados e processados ao longo do tempo. Estudos com dados laboratoriais rotineiros mostram, por exemplo, que mudanças de método podem introduzir descontinuidades artificiais em séries históricas, comprometendo análises longitudinais. De forma semelhante, comparações entre bases derivadas dos mesmos prontuários eletrônicos evidenciam que diferentes processos de extração, transformação e carregamento podem alterar indicadores e resultados analíticos relevantes [3,4].

No plano internacional, a *Global strategy on digital health 2020–2025* da Organização Mundial da Saúde reforça que a transformação digital em saúde depende de estratégias integradas, governança, interoperabilidade e uso seguro, ético e orientado por propósito dos dados. No contexto brasileiro, a Estratégia de Saúde Digital para o Brasil 2020–2028 atribui centralidade à Rede Nacional de Dados em Saúde (RNDS), concebida como plataforma nacional de inovação, informação e serviços digitais em saúde, e explicita prioridades diretamente relacionadas a interoperabilidade, padrões, terminologias, uso secundário dos dados e ecossistema de inovação. Nesse sentido, mapear como a literatura tem tratado o uso secundário de dados em registros eletrônicos de saúde contribui não apenas para a organização do debate internacional, mas também para a compreensão de desafios e oportunidades relevantes para o Sistema Único de Saúde (SUS) e para a consolidação da RNDS [5,6].

Apesar da relevância crescente do tema, a produção científica sobre uso secundário de dados em saúde permanece dispersa entre diferentes finalidades, contextos e abordagens metodológicas. Nesse sentido, uma revisão de escopo pode contribuir para mapear como a literatura recente descreve a reutilização de dados derivados de registros eletrônicos de saúde, quais usos aparecem com maior frequência e de que modo interoperabilidade, qualidade do dado e processamento da informação atravessam esse processo. Assim, o objetivo desta revisão de escopo é mapear como a literatura descreve os usos secundários de dados rotineiramente coletados em RES, com ênfase nas finalidades de reutilização identificadas e no papel da interoperabilidade, da estruturação da informação e da qualidade do dado nesse processo.

2 MÉTODOS

Trata-se de uma revisão de escopo, conduzida com base nas recomendações do *Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews* (PRISMA-ScR) [7]. O protocolo metodológico foi previamente estruturado na plataforma Parsif.al, utilizada para planejamento da revisão, gerenciamento das referências, remoção de duplicatas e apoio à triagem dos estudos. Não houve registro público externo do protocolo.

A revisão foi orientada pela seguinte questão: como a literatura descreve os usos secundários de dados rotineiramente coletados em registros eletrônicos de saúde, com ênfase nas finalidades de reutilização identificadas e no papel da interoperabilidade, da estruturação da informação e da qualidade do dado nesse processo? Foram considerados elegíveis estudos primários e estudos aplicados ou metodológicos que abordassem explicitamente o uso secundário, a reutilização ou o reaproveitamento de dados obtidos de RES, prontuários eletrônicos, *data warehouses* clínicos ou repositórios derivados. Foram incluídos estudos publicados em português ou inglês, no recorte temporal previamente definido de 10 anos. Foram excluídos artigos de revisão, editoriais, comentários, textos de opinião, estudos sem relação clara com reutilização secundária dos dados e trabalhos centrados exclusivamente no uso assistencial primário dos sistemas, sem interface analítica, investigativa, gerencial ou de vigilância.

As buscas foram realizadas nas bases Scopus, PubMed/MEDLINE e Biblioteca Virtual em Saúde (BVS), por serem fontes capazes de recuperar literatura relevante das áreas de Informática em Saúde, Saúde Digital e pesquisa em saúde. A estratégia de busca foi construída a partir de Descritores em Ciências da Saúde (DeCS) relacionados ao tema, complementados por sinônimos, variantes em inglês e termos livres de uso corrente na literatura, sendo adaptada à sintaxe de cada base. Os descritores centrais utilizados corresponderam aos eixos conceituais de dados de saúde coletados rotineiramente, registros eletrônicos de saúde e interoperabilidade da informação em saúde. A estratégia de busca por base é apresentada no Quadro 1. A busca mais recente foi realizada em 29 de abril de 2026, conforme recomendação do PRISMA-ScR.

Os resultados recuperados foram exportados das bases e importados para a plataforma Parsif.al. Após a identificação e a remoção de duplicatas, procedeu-se à seleção em duas etapas: leitura de títulos e resumos, seguida de leitura dos textos completos dos estudos potencialmente elegíveis. A elegibilidade foi definida com base nos critérios previamente estabelecidos. Para fins analíticos, o corpus principal da revisão considerou o recorte temporal de 10 anos adotado no estudo. Trabalhos anteriores a esse intervalo, ainda que conceitualmente aderentes ao tema, foram mantidos apenas como referencial histórico e contextual para subsidiar a introdução e a discussão, não integrando a síntese principal dos resultados.

O processo de *charting* dos dados foi orientado por formulário previamente definido no Parsif.al. Para cada estudo incluído, foram extraídas as seguintes informações: título, autoria, ano de publicação, idioma, país ou região de origem, tipo de estudo, contexto de saúde, tipo de fonte de dados, uso secundário identificado, objetivo, principais achados e observações analíticas do revisor. Também foram registrados elementos relacionados ao papel da interoperabilidade, da qualidade do dado, da estruturação da informação e das estratégias de integração ou processamento dos dados, quando presentes.

A extração dos dados foi realizada por dois revisores independentes, e as divergências foram resolvidas por consenso entre os autores. Em seguida, os dados extraídos foram organizados em um quadro sintético e analisados por síntese narrativa, com ênfase nas finalidades de reutilização identificadas e no papel da interoperabilidade, da qualidade do dado e do processamento da informação. Por se tratar de uma revisão de escopo, sem participação direta de seres humanos nem uso de dados identificáveis, não houve necessidade de submissão ao Comitê de Ética em Pesquisa (CEP), conforme a Resolução nº 510/2016 do Conselho Nacional de Saúde (CNS) [8].

Não foi realizada avaliação crítica formal da qualidade metodológica das fontes incluídas. Essa decisão foi adotada porque o objetivo da revisão foi mapear e caracterizar o campo, identificar tipos de uso secundário e sintetizar abordagens e finalidades descritas na literatura, e não comparar efeitos ou hierarquizar robustez de evidências. Essa opção é compatível com o PRISMA-ScR, que trata a avaliação crítica das fontes de evidência como item condicional, quando aplicável.

Os dados extraídos foram sintetizados por análise descritiva e narrativa. Os estudos incluídos foram agrupados por similaridade temática, com ênfase nas finalidades de uso secundário identificadas e no papel desempenhado pela interoperabilidade, pela qualidade do dado e pelos processos de estruturação, integração e reutilização da informação em saúde. Os resultados são apresentados na seção Resultados por meio da Figura 1, do Quadro 2 e de síntese narrativa dos usos mapeados.

3 RESULTADOS

A estratégia de busca e a organização das referências na plataforma Parsif.al resultaram em 45 registros importados para triagem. Após a identificação de duplicatas, foram reconhecidos 24 registros duplicados, restando 21 registros únicos para análise por título e resumo. Nessa etapa, 7 registros foram excluídos por não atenderem aos critérios de elegibilidade, enquanto 14 estudos foram considerados aderentes ao tema e mantidos para leitura em texto completo. O detalhamento do processo de seleção das fontes de evidência é apresentado na Figura 1, em formato aderente ao PRISMA Flow.

Quadro 1. Estratégia de busca nas bases consultadas

Bases	String	Total de registros importados
Scopus; PubMed/MEDLINE;BVS	(("routinely collected health data" OR "routine health data" OR "routinely collected data" OR "dados de saúde coletados rotineiramente" OR "dados coletados rotineiramente") AND ("electronic health record" OR "electronic health records" OR "electronic medical record" OR "electronic medical records" OR "registro eletrônico de saúde" OR "registros eletrônicos de saúde" OR "prontuário eletrônico" OR "prontuários eletrônicos" OR EHR OR EMR) AND ("health information interoperability" OR interoperability OR "health information exchange" OR "interoperabilidade da informação em saúde" OR "troca de informação em saúde"))	45

Fonte: Elaborado pelos autores (2026).

Entre os 14 estudos considerados elegíveis após leitura de título e resumo, observou-se distribuição temporal concentrada nos anos mais recentes, com parte do material situada fora do recorte principal de 10 anos adotado nesta revisão. Assim, para fins de síntese principal, foram incluídos 9 estudos publicados entre 2018 e 2025, enquanto 5 estudos anteriores a 2017 foram preservados apenas como referencial histórico e contextual. Essa decisão permitiu manter coerência com o recorte temporal definido nos métodos, sem desconsiderar contribuições relevantes para a compreensão da evolução do campo.

Os 9 estudos incluídos no corpus principal apresentaram heterogeneidade de objetivos, cenários e finalidades analíticas, mas mostraram convergência em torno da reutilização de dados provenientes de registros eletrônicos de saúde. De modo geral, os usos secundários identificados puderam ser organizados em cinco eixos principais:

(i) Estruturação, qualificação e preparação dos dados para reutilização;

(ii) Integração multicêntrica e construção de bases analíticas ou de *research-ready datasets*;

(iii) Vigilância, fenotipagem e produção de evidências do mundo real;

(iv) Pesquisa clínica, incluindo reaproveitamento de dados assistenciais em estudos e ensaios; e

(v) Pesquisa em serviços de saúde, qualidade assistencial e gestão hospitalar.

No primeiro eixo, destacaram-se estudos voltados à estruturação, padronização e qualificação dos dados, sobretudo em cenários nos quais a baixa organização da informação limita sua reutilização em pesquisa. Nessa linha, observou-se preocupação com padronização terminológica, aumento da estruturação de prescrições e criação de infraestrutura automatizada para transformar dados assistenciais heterogêneos em dados reutilizáveis e aderentes a princípios de interoperabilidade e FAIRificação [9,10]. No mesmo sentido, um dos estudos mostrou que diferenças nos processos de extração, transformação e carregamento de dados podem alterar indicadores analíticos relevantes, reforçando a importância de metadados, transparência e governança do processamento da informação [4].

No segundo eixo, foram identificadas iniciativas de integração multicêntrica e formação de bases analíticas destinadas à pesquisa clínica e translacional. Esses estudos mostraram a criação de *pipelines* e plataformas capazes de consolidar dados clínicos oriundos de diferentes instituições, com harmonização estrutural e semântica, anonimização, validação e disponibilização para pesquisa colaborativa. Nesses casos, o uso secundário apareceu fortemente vinculado à construção de infraestrutura institucional para compartilhamento, agregação e reutilização de dados em escala ampliada [11,12].

No terceiro eixo, o uso secundário apareceu associado à vigilância, monitoramento e produção de evidências do mundo real. Os estudos incluídos mostraram que dados rotineiramente coletados podem ser reutilizados para identificação de casos, definição de fenótipos computáveis, vigilância sindrômica, monitoramento populacional e avaliação de eventos em larga escala. Nesses trabalhos, a interoperabilidade semântica e o uso de modelos computáveis foram relevantes para tornar critérios clínicos mais compartilháveis e executáveis em diferentes contextos analíticos [12,13].

No quarto eixo, observaram-se usos relacionados à pesquisa clínica, com destaque para reaproveitamento de dados assistenciais em estudos observacionais e ensaios clínicos. Em um dos trabalhos, os dados do prontuário eletrônico foram utilizados como fonte estruturada para auditoria e pesquisa a partir da captura em tempo real no ponto de cuidado. Em outro, a reutilização dos dados foi direcionada à transferência automatizada de informações dos registros eletrônicos para sistemas de coleta de dados de pesquisa, configurando um cenário de uso como *eSource* em ambiente multicêntrico [14,15].

No quinto eixo, os estudos incluídos também revelaram usos secundários voltados à pesquisa em serviços de saúde, qualidade assistencial e gestão hospitalar. Nesses casos, os dados rotineiramente coletados foram empregados para avaliar desempenho organizacional, comportamento de indicadores, suporte a decisões gerenciais e análise de métricas associadas ao cuidado. Esse grupo amplia a compreensão do uso secundário para além da pesquisa acadêmica tradicional, evidenciando sua relevância para avaliação de processos assistenciais e gestão em saúde [4,12,16].

De forma transversal, a interoperabilidade apareceu como elemento estruturante da maior parte dos estudos incluídos. Ela foi mobilizada sob diferentes formas, como padronização terminológica, uso de modelos de informação, integração entre sistemas, harmonização de dados, apoio ao intercâmbio entre bases e explicitação de processos de transformação da informação. Em paralelo, a qualidade do dado, a estruturação da informação e a proveniência surgiram como condições necessárias para que os dados oriundos de RES possam ser reutilizados de forma consistente em pesquisa, vigilância, gestão e inovação em Saúde Digital.

Quadro 2. Características sintéticas dos estudos incluídos no recorte principal

Autor/ano	País/escopo	Uso secundário principal	Elemento técnico central
Reinecke *et al.*, 2023	Alemanha	Estruturação e qualificação de dados de medicamentos	Padronização e aumento da estruturação
Parciak *et al.*, 2023	Alemanha	Integração automatizada e FAIRificação de dados	Integração automatizada, interoperabilidade e FAIRificação
Bodagh *et al.*, 2018	Reino Unido	Reuso de dados assistenciais para pesquisa e auditoria	Registro estruturado em tempo real no ponto de cuidado
van Essen *et al.*, 2025	Países Baixos	Avaliação do impacto do processamento dos dados em resultados epidemiológicos	ETL, proveniência, transparência e governança do dado
Smith *et al.*, 2020	Reino Unido	Consolidação de dados clínicos para pesquisa colaborativa	Harmonização e integração multicêntrica
Jamie *et al.*, 2024	Inglaterra	Fenotipagem computável, vigilância e evidências de mundo real	Interoperabilidade semântica e modelagem computável
Tarabichi *et al.*, 2021	Estados Unidos	Agregação de dados de prontuário eletrônico para pesquisa	Normalização e consolidação entre instituições
Ammour *et al.*, 2023	Europa (multicêntrico)	Transferência automatizada de dados assistenciais para pesquisa	Interoperabilidade entre RES e sistemas de pesquisa
Sapey *et al.*, 2022	Inglaterra	Avaliação assistencial e gestão hospitalar com dados rotineiros	Reutilização analítica de dados assistenciais para apoio à gestão

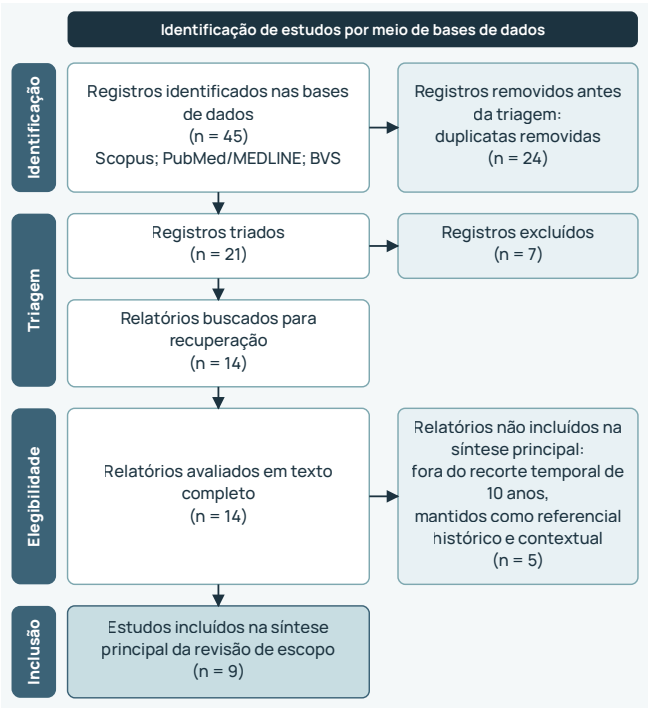
Fonte: Elaborado pelos autores (2026).

Os estudos anteriores ao recorte principal reforçaram, em perspectiva histórica, três núcleos de preocupação já presentes no início da década de 2010: a construção de registros derivados de prontuários eletrônicos, a interoperabilidade semântica voltada à pesquisa clínica e o controle de artefatos metodológicos capazes de comprometer análises com dados rotineiramente coletados. Embora não integrem a síntese principal desta revisão, esses estudos auxiliam na contextualização da trajetória do campo e ajudam a compreender o amadurecimento posterior de temas como FAIRificação, *research-ready datasets*, *eSource* e integração multicêntrica [1-3].

O Quadro 2 apresenta as características sintéticas dos estudos incluídos no recorte principal, com destaque para os usos secundários identificados e os principais elementos técnicos associados à reutilização dos dados:

Os estudos incluídos foram publicados entre 2018 e 2025 e apresentam predominância de desenhos observacionais, estudos de desenvolvimento de infraestrutura e iniciativas multicêntricas voltadas à reutilização de dados provenientes de registros eletrônicos de saúde. Em termos de escopo, observaram-se contribuições sobretudo de contextos europeus e anglófonos, sem identificação de estudos brasileiros ou latino-americanos no corpus principal recuperado pela estratégia desta revisão. Esse achado sugere uma lacuna relevante para a produção científica regional, especialmente diante da centralidade atribuída à interoperabilidade, à governança da informação e ao uso secundário de dados nas agendas de Saúde Digital do SUS, da RNDS e da América Latina.

Figura 1. Fluxograma de seleção das fontes de evidência



Fonte: Elaborado pelos autores (2026), a partir de [7].

4 DISCUSSÃO

Em conjunto, os achados indicam que o uso secundário de dados em RES deve ser compreendido menos como consequência automática da digitalização assistencial e mais como resultado de condições sociotécnicas que tornam os dados disponíveis, estruturados, interpretáveis e adequados a novas finalidades analíticas. Nesse sentido, a revisão desloca o foco da simples existência do prontuário eletrônico para os requisitos de qualidade, interoperabilidade,

proveniência e governança que sustentam sua reutilização em pesquisa, vigilância, gestão e inovação em Saúde Digital [4,9-16].

Um primeiro aspecto que se destaca é que a interoperabilidade aparece de forma recorrente, mas não como fim em si mesma. Nos estudos incluídos, ela surge como componente viabilizador da reutilização secundária, seja por meio da harmonização estrutural e semântica, da integração entre instituições, da modelagem computável de fenótipos, da transferência automatizada de dados entre sistemas ou da construção de infraestruturas voltadas à FAIRificação e ao compartilhamento de dados para pesquisa [10-13,15]. Essa constatação reforça que o uso secundário de dados derivados de RES deve ser analisado em articulação com dimensões como qualidade da informação, padronização terminológica, proveniência e governança de dados. Assim, os achados da revisão indicam que a discussão sobre reutilização secundária não deve ser reduzida à simples digitalização dos processos assistenciais, mas compreendida como parte de uma infraestrutura sociotécnica mais ampla, necessária para tornar os dados recuperáveis, interpretáveis, comparáveis e reutilizáveis em diferentes contextos.

A ausência de estudos brasileiros ou latino-americanos no corpus principal reforça a necessidade de ampliar a produção regional sobre modelos, condições e limites da reutilização secundária de dados em registros eletrônicos de saúde, especialmente em sistemas públicos, universais ou segmentados, nos quais interoperabilidade, governança, sustentabilidade e equidade assumem papel estratégico.

Outro ponto importante é que os estudos mais recentes demonstram um deslocamento do foco exclusivamente descritivo para abordagens mais operacionais e de infraestrutura. Em vez de apenas reconhecer o potencial do prontuário eletrônico como fonte de dados, parte da literatura recente se concentra em como tornar esses dados efetivamente reutilizáveis, com atenção a *pipelines* de integração, processos de extração, transformação e carregamento, transparência metodológica, dados *research-ready* e registro estruturado no ponto de cuidado [4,10,14]. Esse movimento sugere um amadurecimento do campo, no qual o debate sobre uso secundário passa a incorporar exigências mais robustas de reprodutibilidade, rastreabilidade e adequação ao propósito analítico.

Essa dimensão operacional evidencia os bastidores técnicos da reutilização dos dados. A passagem do registro assistencial para o dado analítico não é neutra: escolhas de extração, transformação e carregamento, critérios de inclusão e exclusão, estratégias de mapeamento terminológico, tratamento de duplicidades e decisões sobre completude podem alterar a composição da amostra e os resultados produzidos. Em contextos nos quais anotações clínicas, campos estruturados e práticas de registro coexistem de forma heterogênea, diferenças aparentemente operacionais podem introduzir vieses de seleção, comprometer comparabilidade e reduzir a validade de análises longitudinais. Nessa perspectiva, a reutilização secundária de dados requer não apenas interoperabilidade e disponibilidade, mas também transparência metodológica sobre os processos de preparação do dado, de modo a tornar mais rastreáveis os limites e as implicações analíticas de cada *pipeline* de processamento.

A revisão também mostra que os usos secundários identificados extrapolam a pesquisa acadêmica em sentido estrito. Embora estudos multicêntricos e aplicações voltadas à pesquisa clínica ocupem lugar importante no corpus analisado, também foram encontrados usos associados à vigilância, à avaliação de desempenho assistencial e à gestão hospitalar [13,16]. Isso amplia a relevância prática do tema, sobretudo em contextos de Saúde Digital orientados ao SUS e à gestão pública, nos quais a reutilização de dados pode apoiar monitoramento, planejamento, qualificação do cuidado, avaliação de processos e tomada de decisão baseada em evidências. Nessa perspectiva, o uso secundário do dado deixa de ser apenas uma agenda metodológica da pesquisa em saúde e passa a constituir um componente estratégico da inteligência organizacional em serviços e sistemas de saúde.

Ao mesmo tempo, os achados sugerem cautela diante de uma visão excessivamente otimista da reutilização dos dados. O estudo de van Essen *et al.* evidencia que diferenças nos processos de extração e processamento podem alterar resultados epidemiológicos, o que reforça que o dado secundariamente reutilizado não é neutro nem diretamente equivalente ao fenômeno clínico que pretende representar [4]. De modo semelhante, estudos voltados à estruturação e integração dos dados mostram que a reutilização depende de escolhas prévias de modelagem, terminologia, granularidade, completude e registro, o que pode afetar tanto a validade analítica quanto a comparabilidade entre cenários [9-11,13]. Em outras palavras, o uso secundário de dados de registros eletrônicos de saúde exige mediações técnicas e organizacionais que precisam ser explicitadas, documentadas e discutidas criticamente.

Em perspectiva histórica, os estudos anteriores ao recorte principal mostram que preocupações hoje associadas à governança, FAIRificação, *research-ready datasets* e *eSource* já apareciam no início da década de 2010 em três frentes: construção de registros derivados de prontuários eletrônicos, interoperabilidade semântica para pesquisa clínica e controle de artefatos metodológicos em análises com dados rotineiramente coletados [1-3]. Embora não componham a síntese principal, esses estudos ajudam a evidenciar a continuidade entre problemas iniciais de modelagem e

validação dos dados e agendas mais recentes de integração multicêntrica e reutilização automatizada.

Nesse cenário, a consolidação do uso secundário de dados em saúde depende de arranjos sustentáveis de governança, capazes de compatibilizar interoperabilidade, segurança, ética no uso da informação e viabilidade organizacional ao longo do tempo. Sem essas condições, iniciativas promissoras podem permanecer restritas a experiências pontuais, com baixo potencial de escala e frágil integração às rotinas assistenciais, de gestão e de pesquisa.

Esta revisão apresenta limitações que devem ser consideradas. Em primeiro lugar, a busca foi restrita a três fontes de informação, o que pode ter reduzido a recuperação de estudos relevantes publicados em outras bases. Em segundo lugar, a elegibilidade considerou publicações em português e inglês e o recorte temporal de 10 anos; embora essa delimitação favoreça atualidade e aderência ao objetivo da revisão, pode ter sub-representado contribuições anteriores, literatura regional ou publicações indexadas em outras fontes. Assim, a ausência de estudos brasileiros ou latino-americanos deve ser interpretada como ausência no corpus recuperado por esta estratégia, e não como evidência de inexistência de produção sobre o tema. Em terceiro lugar, não foi realizada avaliação crítica formal da qualidade metodológica das fontes incluídas, opção compatível com o objetivo de mapeamento da revisão de escopo, mas que limita inferências sobre robustez das evidências. Adicionalmente, o processo de extração e síntese dos dados esteve sujeito às limitações inerentes à interpretação dos textos completos e à heterogeneidade dos relatos primários. Por fim, a heterogeneidade dos estudos dificultou comparações mais padronizadas entre desenhos, desfechos e contextos de aplicação.

Apesar dessas limitações, os resultados oferecem uma contribuição útil para a área de Informática em Saúde ao sintetizar como o uso secundário de dados de registros eletrônicos de saúde tem sido operacionalizado e quais elementos técnico-informacionais aparecem como mais relevantes nesse processo. Em especial, a revisão sugere que interoperabilidade, qualidade do dado, estruturação da informação e governança do processamento não devem ser tratados como dimensões periféricas, mas como condições constitutivas da reutilização dos dados. Para a agenda brasileira de Saúde Digital, isso reforça a importância de investir não apenas na ampliação dos sistemas de registro, mas também em estratégias que favoreçam padronização, integração, transparência metodológica e sustentabilidade do uso secundário do dado em benefício da pesquisa, da gestão e do cuidado.

5 CONCLUSÃO

Esta revisão de escopo permitiu mapear como a literatura recente descreve o uso secundário de dados provenientes de registros eletrônicos de saúde, evidenciando que sua reutilização envolve finalidades relacionadas à qualificação dos dados, integração multicêntrica, pesquisa clínica, vigilância, evidências de mundo real e avaliação assistencial e gerencial. Em conjunto, os estudos incluídos indicam que o uso secundário não depende apenas da disponibilidade de registros digitais, mas de condições técnico-informacionais e organizacionais que viabilizem a reutilização consistente dos dados, incluindo estruturação, padronização, interoperabilidade, proveniência, transparência metodológica e governança do processamento.

Ao responder ao objetivo proposto, a revisão sugere que o uso secundário de dados em saúde deve ser compreendido como um processo sociotécnico de transformação dos dados assistenciais em insumos reutilizáveis para pesquisa, monitoramento, gestão e inovação em Saúde Digital. No contexto brasileiro e latino-americano, os achados apontam a necessidade de ampliar investigações que documentem experiências regionais, modelos de governança, infraestruturas de dados e condições de sustentabilidade para reutilização secundária em sistemas de saúde. Estudos futuros podem ampliar essa agenda por meio da inclusão de outras bases, de recortes temáticos complementares e do aprofundamento analítico sobre contextos, métodos e modelos de reutilização dos dados.

Contribuição dos autores. Os autores contribuíram igualmente em todas as etapas de construção do artigo: concepção e delineamento do estudo; coleta, análise e interpretação dos dados; redação e revisão crítica do manuscrito; e aprovação da versão final.

Agradecimentos. Os autores agradecem ao Hospital Universitário Antônio Pedro, à Universidade Federal Fluminense e à Empresa Brasileira de Serviços Hospitalares (Ebserh, atual HU Brasil) pelo apoio e incentivo institucionais à pesquisa, à produção científica e à inovação em Saúde Digital.

Financiamento. Os autores não declararam fonte de financiamento.

Disponibilidade de dados e materiais. Os autores não declararam a disponibilidade dos dados e materiais da pesquisa.

Conflitos de interesse. Os autores não declararam conflitos de interesse.

Declaração de uso de IA. Os autores declaram o uso de inteligência artificial generativa, por meio do ChatGPT, para apoio à organização textual, revisão de redação e aprimoramento da escrita acadêmica do manuscrito. A concepção do estudo, a definição metodológica, a seleção e análise das fontes de evidência, a verificação das informações e a responsabilidade intelectual pelo conteúdo final permaneceram integralmente sob responsabilidade dos autores.

REFERÊNCIAS

1. Ouagne D, Hussain S, Sadou E, Jaulent MC, Daniel C. The Electronic Healthcare Record for Clinical Research (EHR4CR) information model and terminology. Stud Health Technol Inform. 2012;180:534-8. doi: https://doi.org/10.3233/978-1-61499-101-4-534
2. Liaw ST, Taggart J, Yu H. EHR-based disease registries to support integrated care in a health neighbourhood: an ontology-based methodology. Stud Health Technol Inform. 2014;205:171-4. doi: https://doi.org/10.3233/978-1-61499-432-9-171
3. Poh N, McGovern AP, de Lusignan S. Improving the measurement of longitudinal change in renal function: automated detection of changes in laboratory creatinine assay. J Innov Health Inform. 2015;22(2):293-301. doi: https://doi.org/10.14236/jhi.v22i2.81
4. van Essen MHJ, Twickler R, Weesie YM, Arslan IG, Groenhof F, Peters LL, et al. Implications of Data Extraction and Processing of Electronic Health Records for Epidemiological Research: Observational Study. J Med Internet Res. 2025;27:e64628. doi: https://doi.org/10.2196/64628
5. World Health Organization. Global strategy on digital health 2020-2025 [Internet]. Geneva: World Health Organization; 2021 [citado 2026 out 18]. Disponível em: https://www.who.int/publications/i/item/9789240020924
6. Brasil. Ministério da Saúde. Secretaria-Executiva. Departamento de Informática do SUS. Estratégia de Saúde Digital para o Brasil 2020-2028 [Internet]. Brasília: Ministério da Saúde; 2020 [citado 2026 out 18]. Disponível em: https://bvsms.saude.gov.br/bvs/publicacoes/estrategia_saude_digital_Brasil.pdf
7. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. 2018;169(7):467-73. doi: https://doi.org/10.7326/M18-0850
8. Brasil. Conselho Nacional de Saúde. Resolução nº 510, de 7 de abril de 2016 [Internet]. Diário Oficial da União. 2016 maio 24 [citado 2026 out 18]. Disponível em: https://bvsms.saude.gov.br/bvs/saudelegis/cns/2016/res0510_07_04_2016.html
9. Reinecke I, Siebel J, Fuhrmann S, Fischer A, Sedlmayr M, Weidner J, et al. Assessment and Improvement of Drug Data Structuredness From Electronic Health Records: Algorithm Development and Validation. JMIR Med Inform. 2023;11:e40312. doi: https://doi.org/10.2196/40312
10. Parciak M, Suhr M, Schmidt C, Bönisch C, Löhnhardt B, Kesztyüs D, et al. FAIRness through automation: development of an automated medical data integration infrastructure for FAIR health data in a maximum care university hospital. BMC Med Inform Decis Mak. 2023;23(1):94. doi: https://doi.org/10.1186/s12911-023-02195-3
11. Smith DA, Wang T, Freeman O, Crichton C, Salih H, Matthews PC, et al. National Institute for Health Research Health Informatics Collaborative: development of a pipeline to collate electronic clinical data for viral hepatitis research. BMJ Health Care Inform. 2020;27(3):e100145. doi: https://doi.org/10.1136/bmjhci-2020-100145
12. Tarabichi Y, Frees A, Honeywell S, Huang C, Naidech AM, Moore JH, et al. The Cosmos Collaborative: A Vendor-Facilitated Electronic Health Record Data Aggregation Platform. ACI open. 2021;5(1):e36-e46. doi: https://doi.org/10.1055/s-0041-1731004
13. Jamie G, Elson W, Kar D, Wimalaratna R, Hoang U, Meza-Torres B, et al. Phenotype execution and modeling architecture to support disease surveillance and real-world evidence studies: English sentinel network evaluation. JAMIA Open. 2024;7(2):ooae034. doi: https://doi.org/10.1093/jamiaopen/ooae034
14. Bodagh N, Archbold RA, Weerackody R, Hawking MKD, Barnes MR, Lee AM, et al. Feasibility of real-time capture of routine clinical data in the electronic health record: a hospital-based, observational service-evaluation study. BMJ Open. 2018;8(3):e019790. doi: https://doi.org/10.1136/bmjopen-2017-019790
15. Ammour N, Griffon N, Djadi-Prat J, Chatellier G, Lewi M, Todorovic M, et al. TransFAIR study: a European multicentre experimental comparison of EHR2EDC technology to the usual manual method for eCRF data collection. BMJ Health Care Inform. 2023;30(1):e100602. doi: https://doi.org/10.1136/bmjhci-2022-100602
16. Sapey E, Gallier S, Evison F, McNulty D, Reeves K, Ball S. Variability and performance of NHS England's 'reason to reside' criteria in predicting hospital discharge in acute hospitals in England: a retrospective, observational cohort study. BMJ Open. 2022;12(12):e065862. doi: https://doi.org/10.1136/bmjopen-2022-065862

NOTAS EDITORIAIS

O Journal of Health Informatics é uma publicação científica mantida pela Sociedade Brasileira de Informática em Saúde, associação científica sem fins lucrativos dedicada ao desenvolvimento da Informática em Saúde e da Saúde Digital no Brasil. O processo editorial é conduzido com independência científica, sem interferência institucional, comercial ou de patrocinadores, e segue as boas práticas internacionais de ética em publicação científica.

Este artigo foi avaliado por pares, em sistema duplo-cego, por especialistas da área. A equipe editorial usou ferramentas de inteligência artificial, com supervisão humana, na triagem inicial, na conferência do texto, das referências e dos números e, quando necessário, no redesenho de figuras, quadros e tabelas, aprovado pelos autores na prova. As ferramentas e as salvaguardas estão na Política de uso de inteligência artificial da revista. As decisões editoriais foram tomadas por pessoas.

SBIS Sociedade Brasileira de Informática em Saúde