

# Dados administrativos brasileiros para pesquisa do mundo real: Método de Linkage Determinística e Harmonização OMOP CDM

**Autores:** Thales Brendon Castano Silva, Jessica Mayumi Maruyama, Julio Cesar Barbour Oliveira

**Instituição:** Precision Data Engineering & Data Science - São Paulo - SP - Brasil

**Introdução:** Evidência do mundo real (RWE) fornece informações valiosas sobre a segurança, a eficácia e o valor das intervenções de saúde em populações de pacientes do mundo real e permite o exame das jornadas dos pacientes. As informações coletadas pelo DATASUS não são integradas pelos diferentes sistemas e os dados coletados por diferentes serviços de saúde carecem de identificadores-chave únicos no nível individual, o que representa um desafio para o relacionamento de registros de tais conjuntos de dados. **Objetivos:** Descrever os métodos e os resultados parciais do produto mínimo viável da definição de parâmetros necessários para criar um conjunto de dados usando dados de sinistros brasileiros, avaliando a qualidade dos dados em relação a uma instância OMOP CDM usando o DataQualityDashboard. **Material e Método:** As bases de dados do DATASUS utilizadas foram o Sistema de Informação Hospitalar e o Ambulatorial de janeiro de 2008 até a presente data. Atualmente, estamos desenvolvendo um pipeline de dados, que passará por atualizações mensais. As informações disponíveis em ambos os conjuntos de dados incluem todos os procedimentos ambulatoriais, consultas, códigos CID-10 de diagnóstico primário e secundário, medicamentos dispensados para uso domiciliar e dados pessoais, como data de nascimento, endereço residencial, CEP e sexo. Todos os registros originais do banco de dados DATASUS foram analisados para avaliar a consistência das informações para uma chave única de paciente. Após a etapa de limpeza e pré-processamento, um algoritmo de ligação determinística foi desenvolvido para conectar hospitais com registros ambulatoriais usando as informações-chave de CEP, data de nascimento e sexo. O processo de padronização no processo OMOP seguiu as diretrizes fornecidas pelos manuais da comunidade online e a verificação da qualidade dos dados foi realizada usando o DataQualityDashboard versão 2.1.1.7. **Resultados:** Foi construído um conjunto de dados padronizado que abrange o histórico completo de saúde desses 5,82 milhões de pacientes. Dessas 2.165 verificações que avaliam a plausibilidade dos dados, apenas 24 falharam em nosso banco de dados, enquanto 2.141 foram aprovadas. No mapeamento de qualidade da base de dados, encontramos 88 falhas e passamos com sucesso 3.578 consultas. Nosso banco de dados apresentou uma taxa de aprovação geral de 98%, indicando um índice satisfatório de qualidade do mapeamento. **Discussão e Conclusões:** Os resultados do presente estudo demonstraram uma metodologia de tratamento de dados para conjuntos de dados do DATASUS com forte qualidade e o fornecimento de um novo banco de dados que facilita a colaboração dentro da comunidade OHDSI. Através da harmonização dos dados, facilitamos a colaboração e a troca de informações entre pesquisadores, instituições e setores da indústria.

**Palavras-Chave:** Real-world evidence; Real-world data; DATASUS.

## Referências Bibliográficas:

1. Liu F, Panagiotakos D. Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC Med Res Methodol* 2022; 22:287 <https://doi.org/10.1186/s12874-022-01768-6>.
2. Nabhan C, Klink A, Prasad V. Real-world evidence - what does it really mean? *JAMA Oncol*. 2019;5(6):781.doi: 10.1001/jamaoncol.2019.0450
3. Justo N, Espinoza MA, Ratto B, et al. Real-world evidence in healthcare decision making: global trends and case studies from latin america. *Value Health*. 2019;22(6):739-749.
4. Ali MS, Ichihara MY, Lopes LC, Barbosa GCG, Pita R, Carreiro RP, Dos Santos DB, Ramos D, Bispo N, Raynal F, Canuto V, de Araujo Almeida B, Fiaccone RL, Barreto ME, Smeeth L, Barreto ML. Administrative data linkage in Brazil: potentials for health technology assessment. *Front Pharmacol*. 2019;10:984. doi: 10.3389/fphar.2019.00984.
5. Junior AAG, Pereira RG, Gurgel EI, Cherchiglia M, Dias LV, Ávila J, et al. Building the national database of health centred on the individual: administrative and epidemiological record linkage - Brazil, 2000-2015. *Int J Popul Data Sci*. 2018; 3(1). doi: 10.23889/ijpds.v3i1.446.
6. Junior EPP, Normando P, Flores-Ortiz R, et al. Integrating real-world data from Brazil and Pakistan into the OMOP common data model and standardized health analytics framework to characterize COVID-19 in the Global South. *J Am Med Inform Assoc*. 2023; 30(4):643-655. doi:10.1093/jamia/ocac180
7. DATASUS – Ministério da Saúde [Internet]. [Brasília] Ministério da Saúde (BR). [cited 2023 May 30]. Available from: <https://datasus.saude.gov.br/>  
Blacketer C, Schuemie FJ, Ryan PB, Rijnbeek P. "Increasing trust in real-world evidence through evaluation of observational data quality. *J Am Med Inform Assoc* 2021;28(10):2251-2257. <https://doi.org/10.1093/jamia/ocab132>.