

International Journal of Population Data Science

Journal Website: www.ijpds.org



Methodological Challenges when Using Routinely Collected Health Data for Research: A scoping review.

Lili Wei¹, Ellen Kuenzig¹, James Im¹, Yan Zheng², Taylor McLinden³, Scott Emerson⁴, Azza Eissa⁵, Henry Halder⁶, Richard Shaw⁷, An-Wen Chan⁵, William Dixon⁸, Vera Ehrenstein⁹, Astrid Guttman⁵, Katie Harron¹⁰, Lars G. Hemkens¹¹, Asbjørn Hróbjartsson¹², Ronan Lyons¹³, Shannon E. MacDonald¹⁴, Jerry Maniate¹⁵, David Moher¹⁵, Irene Petersen¹⁰, Hude Quan¹⁶, Sigrún Schmidt⁹, Henrik Sørensen⁹, Shirley Wang¹⁷, David A. McAllister⁷, Sinéad M. Langan¹⁸, and Eric I. Benchimol¹

¹The Hospital for Sick Children

²Tel Aviv University

³Simon Fraser University

⁴BC Centre for Excellence in HIV/AIDS

⁵University of Toronto

⁶Alberta Health Services

⁷University of Glasgow

⁸University of Manchester

⁹Aarhus University

¹⁰University College London

¹¹University Hospital Basel

¹²Odense University Hospital

¹³Swansea University

¹⁴University of Alberta

¹⁵University of Ottawa

¹⁶University of Calgary

¹⁷Harvard University

¹⁸The London School of Hygiene & Tropical Medicine

Routinely collected health data (RCD) including electronic health records, disease registries, health administrative data and wearables data are not specifically collected for research purposes. Analysis of these data poses unique methodological challenges that must be addressed when conducting research, particularly as availability and use increase.

This scoping review aimed to identify methodological challenges in research using RCD from existing literature (registered protocol: <https://doi.org/10.17605/OSF.IO/EBM4D>). We searched 6 electronic databases, including medical, health economics, nursing and psychology research databases, between Jan 2015 and Jan 2023, combining multiple “RCD” and “research” search terms (e.g., epidemiologic, informatics, pharmaceutical research). After screening abstracts and full-texts, we doubly extracted methodological themes, categorizing them into different study stages.

We screened more than 23,000 records and included 430 papers. Bias and confounding were the most common methodological issues identified, discussed in relation to both study design and data analysis. Data quality, including data accuracy, validation, completeness, timeliness and cleaning, also posed substantial challenges, particularly during data processing stage. Record linkage and conducting analyses using dis-

tributed health networks also pose unique methodological challenges. Heterogeneity, incorporating social determinants of health and statistical models that address methodological challenges are also described in the literature. External validity and reporting are important considerations for RCD research.

Our review identified several methodological challenges facing researchers using RCD. These issues should be addressed to ensure methodologically sound research. These findings will inform the development of a standardized protocol template and accompanying educational platform aimed at enhancing methodological quality and transparency when conducting research using RCD

