

International Journal of Population Data Science

Journal Website: www.ijpds.org



The Impact of Data Linkage Errors on Health and Social Research: A Systematic Scoping Review

Barret A. Monchka¹, Muditha Lakmali Bodawatte Gedara², Razieh Dorostimotlagh², Marcelo Urquia², Sumeet Kalia², and Lisa M. Lix¹

¹University of Manitoba, Winnipeg, Canada; George & Fay Yee Centre for Healthcare Innovation, Winnipeg, Canada

²University of Manitoba, Winnipeg, Canada

Background

Record linkage enables research using data from multiple sources but can introduce linkage errors—missed matches and false matches—that may bias downstream analyses. Evidence on their impact has not been systematically synthesized to inform linkage practice or statistical adjustment. Objective: To synthesize published evidence on the impact of record linkage errors on downstream analyses in health and social science research.

and transparent reporting practices to support assessment of linkage quality and potential sources of bias.

Approach

A comprehensive literature search was conducted in Embase. Reference list scanning and supplemental hand searching were used to identify additional relevant studies not captured in the systematic search.

Results

The systematic search identified 228 articles. Title and abstract screening yielded 47 articles for full-text review, of which seven met the inclusion criteria. Two additional articles were identified through reference list scanning and four through hand searching, resulting in 13 included publications. Eleven studies (85%) reported that linkage errors had a considerable impact on study findings, specifically on Cox proportional hazards ratios ($n = 5$), logistic regression odds ratios ($n = 2$), and rates or ratios ($n = 4$). False matches were shown to reduce estimated effect sizes, whereas missed matches increased standard errors, reducing the precision of estimates, and introduced sampling error by disproportionately excluding certain population subgroups.

Conclusions

Evidence suggests that data linkage errors can substantially affect research findings across analytical methods. These findings highlight the importance of accurate linkage processes, appropriate statistical adjustment for residual linkage error,

