

International Journal of Population Data Science

Journal Website: www.ijpds.org

Using the RECORD guidelines to improve transparent reporting of studies based on routinely collected data

Harron, K^{1*}, Benchimol, E^{2,3}, and Langan, S⁴

Submission History

Submitted:	14/09/2017
Accepted:	14/12/2017
Published:	24/01/2018

¹Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine

²Department of Pediatrics and School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, Ottawa, Canada

³Institute for Clinical Evaluative Sciences, Ottawa, Canada

⁴Department of Non-communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine

Summary

Transparent reporting of routinely-collected data studies is key to producing valid and reliable research that can inform decisions about patient care and health systems. This article discusses some of the unique challenges in using these data sources, and explains how the REporting of studies Conducted using Observational Routinely-collected Data (RECORD) guidelines were developed to help researchers and journals to maintain a high level of quality in reporting of healthcare studies using routinely-collected data.

Routinely-collected data, i.e. individual-level records collected for financial, administrative, or clinical management purposes, contain rich, detailed information on patient pathways, and their great potential for research and quality improvement has been increasingly exploited internationally over recent years¹. The strengths of routinely-collected data and their value in health research are well established. However, researchers are faced with unique challenges when conducting research with data sources that were collected for other purposes². In particular, data quality is an ongoing concern - electronic clinical data do not always contain the complete, accurate information that researchers require. For example, consider a simple study comparing patients with and without diabetes. If a patient is to be correctly classified as having diabetes, we require that i) the clinician recognises the diagnosis, ii) the diagnosis is recorded in clinical notes, iii) the medical coders correctly code the diagnosis and iv) the researcher includes the correct codes in their analysis. Omissions in any of these steps could lead to missing information. Missing data can lead to bias - smoking and blood pressure may be important confounders for diabetes, yet meaningful smoking status is rarely available in electronic clinical data unless there are incentives for recording, and missingness in blood pressure measurements may be informative (e.g. data points are more likely to be recorded in patients with high blood pressure values)³. Completeness of data may also be related to external factors such as financial incentives to code specific

indicators or diseases⁴. The limitations of electronic clinical data and inherent challenges for analysis are well recognised by researchers⁵. Transparency in reporting of these issues is key to producing valid and reliable research, allowing thoughtful interpretation of findings, and enabling those working in policy to make informed decisions on patient care and health systems based on electronic data sources.

Reporting guidelines such as CONSORT for clinical trials and STROBE for observational studies aim to improve transparency, allowing identification of potential biases, critical assessments of robustness, and importantly, replication in different settings^{6, 7}. Many leading health journals now actively endorse reporting guidelines, requiring authors to submit relevant checklists or referring authors to the EQUATOR (Enhancing the QUALity and Transparency Of health Research) network website (www.equator-network.org) in their Instructions to Authors. The EQUATOR network was established to improve the reliability and usability of health research literature by facilitating accurate and complete reporting of research studies⁸. Their website provides a comprehensive collection of resources and reporting guidelines to support transparent publication of research.

Until recently, no guidelines adequately addressed the unique challenges of using routinely-collected data for research. To fill this gap, we proposed an extension to the STROBE checklist, the guidelines for the REporting of studies Conducted using Observational Routinely-collected Data

*Corresponding Author:

Email Address: k.harron@ucl.ac.uk (K Harron)

(RECORD)⁹. The RECORD initiative involved Delphi surveys of over 100 international stakeholders to prioritize themes, face-to-face meetings of a working committee to establish consensus and wording, and feedback and review by stakeholders before the final checklist and explanatory document were produced¹⁰. RECORD aims to provide guidance for researchers, peer reviewers, journal editors and readers of studies based on routinely-collected data. Our website (record-statement.org) provides access to the RECORD checklist and an explanatory document with examples of studies that report in a transparent way. Translations are currently available in Chinese, Japanese and German, and further translations are being developed. The website also contains a list of journals that have endorsed and implemented RECORD for submitted manuscripts (including the International Journal of Population Data Science). RECORD is particularly relevant for research articles submitted to the IJPDS, as it highlights specific issues relating to the use of routinely-collected data, including the use of linkage between data sources, access and availability of data, and validation of codes or algorithms used to identify subjects, exposures, outcomes or confounders.

Whilst providing robust evidence on the impact of guidelines on quality of reporting is not straightforward, it is clear that these guidelines are a useful tool to help researchers generate accurate and complete representations of their work¹¹⁻¹³. However, due to the way in which routinely-collected data are generated, extracted, and curated for analysis, with fragmented processes often involving multiple organisations, researchers can find it challenging to obtain the information they need to fully report their research. This is particularly relevant for data linkage studies for which trusted third parties are used to link data from different sources. Datasets transferred to the researcher for analysis are often limited in terms of metadata or information about the linkage process, making it difficult for researchers to adhere to reporting guidelines¹⁴. To help address this issue, GUILD (Guidance for Information about Linking Datasets) recommends information that should be made available at each step of the data linkage pathway (by data providers, linkers, analysts and those writing reports)¹⁵. Sharing of this information, whilst preserving data privacy, could improve reproducibility of research, and promote the increased use of methods to address linkage error¹⁶.

As the use of population-level patient data continues to expand, public perception on how and why these data are collected and used is becoming increasingly important, with the media and the public being vocal about the need to understand the purposes for which data are collected and used¹⁷. The impetus is on authors to maintain a high level of quality in reporting of health research, yet journals also have an important role to play. Through encouraging responsible and complete reporting of research using these data, journal endorsement of the RECORD checklist is an important step towards achieving this goal, and we are pleased that the editors of IJPDS support this initiative. However, it is the joint responsibility of journals, researchers and data providers to strive for a high level of transparency, and to support the public in making informed decisions about the use of their data, in order to continue exploiting electronic health data for improving health and health services¹⁸.

References

1. Jutte DP, Roos L, Brownell MD. Administrative record linkage as a tool for public health research. *Annu Rev Public Health*. 2011. 32:91-108 <https://doi.org/10.1146/annurev-publhealth-031210-100700>
2. Hashimoto RE, Brodt ED, Skelly AC, Dettori JR. Administrative database studies: goldmine or goose chase? *Evid Based Spine Care J*. 2014. 5(2):74-6 <https://doi.org/10.1055/s-0034-1390027>
3. Welch C, Bartlett J, Petersen I. Application of multiple imputation using the two-fold fully conditional specification algorithm in longitudinal clinical data. *Stata J*. 2013.
4. Smith CJP, Gribbin J, Challen KB, Hubbard RB. The impact of the 2004 NICE guideline and 2003 General Medical Services contract on COPD in primary care in the UK. *QJM*. 2008. 101(2):145-53 <https://doi.org/10.1093/qjmed/hcm155>
5. van Walraven C, Austin P. Administrative database research has unique characteristics that can risk biased results. *J Clin Epidemiol*. 2012. 65(2):126-31 <https://doi.org/10.1016/j.jclinepi.2011.08.002>
6. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 2001. 134 <https://doi.org/10.7326/0003-4819-134-8-200104170-00012>
7. Elm Ev, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ*. 2007. 335(7624):806-8 <https://doi.org/10.1136/bmj.39335.541782.AD>
8. Simera I. Chapter 5.6: Reporting guidelines: a tool to increase completeness, transparency, and value of health research published in your journal. In: Smart P, Maisonneuve H, Polderman A, editors. *Science Editors' Handbook European Association of Science Editors*. www.ease.org.uk 2013.
9. Nicholls SG, Quach P, von Elm E, Guttman A, Moher D, Petersen I, et al. The REporting of Studies Conducted Using Observational Routinely-Collected Health Data (RECORD) Statement: Methods for Arriving at Consensus and Developing Reporting Guidelines. *PLoS ONE*. 2015. 10(5):e0125620 <https://doi.org/10.1371/journal.pone.0125620>
10. The Plos Medicine Editors. From checklists to tools: Lowering the barrier to better research reporting. *PLoS Med*. 2015. 12(11):e1001910 <https://doi.org/10.1371/journal.pmed.1001910>
11. Turner L, Shamseer L, Altman DG, Weeks L, Peters J, Kober T, et al. Does the CONSORT Statement influence the quality of reporting of RCTs published in medical journals? *Cochrane Database Syst Rev*. 2012. 1(60)

12. Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA*. 2001. 285(15):1992-5 <https://doi.org/10.1001/jama.285.15.1992>
13. Harron K, Wade A, Muller-Pebody B, Goldstein H, Gilbert R. Opening the black box of record linkage. *J Epidemiol Commun H*. 2012. 66(12):1198 <https://doi.org/10.1136/jech-2012-201376>
14. Gilbert R, Lafferty R, Hagger-Johnson G, Harron K, Zhang L-C, Smith P, et al. GUILD: Guidance for Information about Linking Datasets. *J Public Health*. 2017. 1-8 <https://doi.org/10.1093/pubmed/fox037>
15. Harron K, Doidge J, Knight H, Gilbert R, Goldstein H, Cromwell D, et al. A guide to evaluating linkage quality for the analysis of linked data. *Int J Epidemiol*. 2017. 46(5):1699-710 <https://doi.org/10.1093/ije/dyx177>
16. Hoeksma J. The NHS's care.data scheme: what are the risks to privacy? *BMJ*. 2014. 348 <https://doi.org/10.1136/bmj.g1547>
17. McGrail KM, Gutteridge K, Meagher NL. Building on Principles: The Case for Comprehensive, Proportionate Governance of Data Access. In: Gkoulalas-Divanis A, Loukides G, editors. *Medical Data Privacy Handbook*. Cham: Springer International Publishing; 2015. p. 737-64. https://doi.org/10.1007/978-3-319-23633-9_28

References

KH is funded by the Wellcome Trust, grant number 103975/Z/14/Z. The RECORD initiative is funded by the Canadian Institutes of Health Research (grant number 130512), the Swiss National Science Foundation (grant number IZ32Z0_147388/1), and the Aarhus University Department of Clinical Epidemiology.

