

Reporting of studies based on the analysis of routinely collected health data

Clinical data are routinely collected as part of patient care or for administrative and funding purposes. This information is often stored digitally and is therefore relatively easy to analyse. Studies based on these analyses may provide useful insights to complement formal clinical trials and contribute to our scientific knowledge base.^[1] However, these are observational data and prone to various errors and biases, which may undermine the usefulness and generalisability of any information produced. The size of the database does not guarantee reliability – even extremely large datasets, so-called ‘big data’, are at risk.^[2]

Observational data can reflect the ‘real world’. For example, the efficacy and toxicity of a drug are initially evaluated in the setting of a clinical trial. Observational data may be useful in larger and more diverse populations of daily practice. These data can give a better sense of the overall effectiveness of a drug and also lead to the identification of rare adverse effects that may not be seen in a randomised trial. Observational studies may also reveal associations and possible links between diseases and aetiological or prognostic factors, which might be explored and defined in more planned and targeted studies.

However, observational studies based on routine data are more prone to error and bias than planned prospective observational studies. Routine datasets would have been designed for a specific purpose and might not include key variables, which would be important for a prognostic or comparative study but are not relevant to the primary purpose of the dataset. Also, they may not have undergone the same verification and quality assurance as formal planned clinical studies and trials and perhaps be more at risk of coding and transcription errors.

Observational data will also be susceptible to bias, which may be technical or cognitive in nature.^[3] An example of technical bias is when the reason for selecting a patient for an intervention is unclear and may depend on unknown or unrecorded prognostic factors. An example of cognitive bias is the tendency to look for results that confirm our prior beliefs. Large datasets cannot be assumed to be less susceptible to bias than smaller sets.^[2]

Against this background, an international group of more than 100 scientists and stakeholders produced a statement and a checklist on reporting of routinely collected clinical data. The statement is referred to as the RECORD statement (REporting of studies Conducted using Observational Routinely-collected health Data) (<https://www.record-statement.org/checklist.php>).^[4]

The RECORD statement is an extension of a previous statement for using observational data in epidemiological studies, called STROBE (Strengthening the Reporting of Observational Studies in Epidemiology).^[5] This statement has been widely used and endorsed by medical journals.

The RECORD checklist contains items on the title, abstract, introduction, methods, results and discussion in manuscripts. Although all the items on the checklist need to be noted, selected key items are the following:

Introduction. Item 3: state specific study objectives, including any prespecified hypotheses. The authors should clarify whether the study was exploratory and aimed at finding new associations in the data or aimed at confirming whether the data were compatible with

existing hypotheses. It is not sufficient to simply refer to the data as descriptive.

Methods. Item 4: the key elements of study design should be presented. Item 7: describe the variables that might have modified the observed effects. Item 9: describe how potential bias was addressed. Item 12: describe the statistical methods used and how missing data were addressed.

Results. Item 13.1: describe the selection of persons included in the study, including patients excluded because of poor data quality and data availability.

Discussion. Item 19: discuss limitations of the study, considering potential bias or possible inaccurate data. Item 19.1: discuss the implications of using data that were not created or collected to answer the specific research questions. This would include unmeasured confounding factors, missing data and changing eligibility over time.

While adherence to the RECORD statement and checklist offers useful insights, observational studies from routinely collected data have inherent scientific limitations. These will need to be considered within the context of formal clinical studies and other reporting guidelines.^[6]

An important limitation is that economically and socially vulnerable groups may be under-represented in observational studies based on the database population.

The effectiveness of interventions also cannot be reliably deduced from observational data alone, unless the effect is large, it has a clear mechanistic explanation and the interval between intervention and outcome is short.^[7] Nonetheless, there may be important confounders that outweigh the effect of interventions. For example, the long survival of patients with early-stage prostate cancer may depend more on natural history than on any intervention.

The RECORD statement and checklist should be used for the publication of observational studies using routinely collected observational data.

R P Abratt

Icon Oncology, and Division of Radiation Oncology, Department of Radiation Medicine, Faculty of Health Sciences, University of Cape Town, South Africa
raymond.abratt@iconsa.co.za



F Macbeth

Centre for Trials Research, College of Biomedical and Life Sciences, Cardiff University, UK



1. Elliott JH, Grimshaw J, Altman R, et al. Make sense of health data. *Nature* 2015;527(7576):31-32. <https://doi.org/10.1038/527031a>

2. Kaplan RM, Chambers DA, Glasgow RE. Big data and large sample size: A cautionary note on the potential for bias. *Clin Transl Sci* 2014;7(4):342-346. <https://doi.org/10.1111/cts.12178>
3. Macbeth F, Webster A. Understanding bias in the medical literature: With reflections on metastasectomy. *S Afr J Oncol* 2020;4(40):a144. <https://doi.org/10.4102/sajo.v4i0.144>
4. Benchimo E, Smeeth L, Guttman A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015;12(10):e1001885. <https://doi.org/10.1371/journal.pmed.1001885>
5. Vandembroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational studies in Epidemiology (STROBE): Explanation and elaboration. *PLoS Med* 2007;4(10):e297. <https://doi.org/10.1371/journal.pmed.0040297>
6. Abratt RP. How to publish your work in a peer-reviewed journal: A short guide. *S Afr Med J* 2020;110(7):640-641. <https://doi.org/10.7196/SAMJ.2020.v110i7.14679>
7. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 2007;334(7589):349. <https://doi.org/10.1136/bmj.39070.527986.68>

S Afr Med J 2020;110(12):1146-1147. <https://doi.org/10.7196/SAMJ.2020.v110i12.15370>