

Systematic Reviews and Meta-Analyses of Non-randomised Studies



Sanjay Patole

Abstract Randomised controlled trials (RCTs) are considered as the gold standard for clinical research because unlike other study designs, they control for known, and importantly, unknown confounders by randomisation. Evaluation of interventions should hence be ideally done by RCTs. However, RCTs are not always possible or feasible for various reasons, including ethical concerns and the need for time, effort, and funding. Difficulty in the generalisation of the findings of RCTs is also an issue given their rigid design. Non-randomised studies (non-RCTs) provide an alternative to RCTs in such situations. These include cohort, case-control and cross-sectional studies. Non-RCTs have the advantage of providing data from the real-life situation rather than that from the rigid framework of RCTs. The limitations of non-RCTs include selection bias and lack of randomisation that allow confounders to influence the results. At best, non-RCTs can only generate hypotheses for testing in RCTs. This chapter covers the methodology for conducting, reporting and interpreting systematic reviews and meta-analysis of non-RCTs.

Keywords Confounding · MOOSE guidelines · New castle ottawa scale · Non-randomised studies · Randomised controlled trials · Risk of bias · ROBINS-1 tool

Introduction

Randomised controlled trials (RCTs) are considered as the gold standard for clinical research because unlike other study designs, they control for known, and importantly, unknown confounders by randomisation. Allocation concealment protects randomisation. The core elements of the RCT (randomisation, allocation conceal-

S. Patole (✉)

School of Medicine, University of Western Australia, Perth, WA 6009, Australia

e-mail: sanjay.patole@health.wa.gov.au

S. Patole

Neonatal Directorate, King Edward Memorial Hospital for Women,

Perth, WA 6008, Australia

© Springer Nature Switzerland AG 2021

S. Patole (ed.), *Principles and Practice of Systematic Reviews and Meta-Analysis*,

https://doi.org/10.1007/978-3-030-71921-0_13

ment and blinding) minimise bias and optimise the internal validity of the results. Evaluation of interventions should hence be ideally done by RCTs. However, RCTs are not always possible or feasible for various reasons, including ethical issues, and importantly, the need for time, effort, and funding. Definitive trials particularly need significant resources considering their large sample sizes, complexity, logistics and the need for expertise in various aspects of the trial. Difficulty in generalisation (i.e. external validity) of the findings of RCTs with rigid designs is also an issue. Non-randomised studies (non-RCTs) provide an alternative to RCTs in such situations (Mariani and Pego-Fernandes 2014; Gershon et al. 2018; Gilmartin Thomas and Liew 2018; Heikinheimo et al. 2017; Ligthelm et al. 2007; Jepsen et al. 2004). These include cohort (Prospective or retrospective), case-control and cross-sectional studies. Non-RCTs have the advantage of providing data from the real-life situation rather than that from the rigid framework of RCTs.

Cohort studies allow estimation of the relative risk as well as the incidence and natural history of the condition under study. They can differentiate cause from an effect as they measure events in temporal sequence. When designed well, adequately powered prospective cohort studies provide the second-best option after RCT (Mann 2003). Both designs include two groups of participants and assess desired outcomes after exposure to intervention over a specified time in a setting (The PICOS approach). However, the critical difference is that unlike the RCT, the two groups (exposed vs not exposed) are not selected randomly in a cohort study. Retrospective cohorts are quick and cheaper to conduct, but the validity of their results is questionable considering the unreliable and often, inadequate retrospective data.

Unlike cohort studies that can assess common conditions and common exposures, case-control studies help in studying rare conditions/diseases and rare exposures (e.g. lung cancer after asbestos exposure). To put it simply, case-control studies assess the frequency of exposure in those with vs those without the condition/disease of interest. If the frequency of exposure is higher in those with the condition of interest than those without the condition; thus establishing an 'association'. Hill's criteria for associations are important in this context. Case-control studies estimate odds ratios (OR) rather than relative risk (RR). The difficulties in matching control groups for known confounders and a higher risk of bias are limitations of case-control studies. Cross-sectional studies are also relatively quick and cheap, can be used to estimate prevalence, and study multiple outcomes. However, they also cannot differentiate between cause and effect.

Overall, the major limitations of non-randomised studies include selection bias and lack of randomisation that allow confounders to influence the results (Gueyffier and Cucherat 2019; Gerstein et al. 2019). A confounder is any factor related to the intervention as well as the outcome and could affect both. Therefore, at best, non-randomised studies can only generate hypotheses that need to be tested in RCTs. They are useful for identifying associations that can then be more rigorously studied using a cohort study or ideally in an RCT. One of the commonly used statistical tools to address the issue of confounding is regression analysis which 'adjusts/controls' the results for known confounders. This is the reason why access to both, unadjusted as well as adjusted results (e.g. ORs), is important for interpreting the results of non-RCTs. Other techniques such as propensity scores and

sensitivity analysis can reduce bias caused by the lack of randomisation in non-RCTs (Joffe and Rosenbaum 1999). Non-RCTs are known to overestimate the effects of an intervention. However, adequately powered, and well designed and conducted non-RCTs can provide effects estimates that are relatively close to those provided by RCT (Concato et al. 2000).

Despite their limitations, non-RCTs have a substantial and well-defined role in evidence-based practice. They are a crucial part of the knowledge cycle and complement RCTs (Faraoni and Schaefer 2016; Schillaci et al. 2013; Norris et al. 2010). Systematic reviews and meta-analyses of non-RCTs are hence common in all faculties of medicine. This section briefly covers the critical aspects of the process of systematic review and meta-analysis of non-RCTs compared with RCTs.

Conducting a Systematic Review of Non-RCTs

The initial steps in conducting a systematic review of non-RCTs are similar to those for a systematic review of RCTs. These include framing a clinically useful and answerable question using the PICO approach, deciding the type of studies to be searched (e.g. non-RCTs of an intervention), and conducting a comprehensive literature search for the best available evidence. The search is much broader compared to that for RCTs given the different study designs that come under the term “non-RCTs”. To avoid wastage of resources and duplication, it is essential to check whether the question has already been answered.

The search strategy includes the following terms for the publication type: *observational, cohort, case-control, cross-sectional studies, retrospective, prospective studies, non-randomised controlled trial*. Searching major databases, grey literature, proceedings of the relevant conference proceedings, registries, checking cross-references of important publications including reviews, and contacting experts in the field is as important as in any other systematic review.

Having a team of subject experts and methodologists optimises the validity of the results. A transparent and unbiased approach, and use robust methods, and explicit criteria are critical to assure that the review is ‘truly’ systematic (Transparent, Robust, Reproducible, Unbiased, Explicit).

The Cochrane methodology and MOOSE guidelines (Meta-analysis of Observational Studies in Epidemiology) are commonly followed for conducting and reporting systematic reviews of non-RCTs (Lefebvre et al. 2008; Stroup et al. 2000; Lefebvre et al. 2013).

Data Extraction

Data extraction is done independently by at least two reviewers, using the data collection form designed for the review. For dichotomous outcomes, the number of participants with the event and the number analysed in each intervention group of

each study are recorded. Availability of these data helps in creating forest plots of unadjusted ORs.

For continuous outcomes, the mean and standard deviation are entered. Authors of the included studies may need to be contacted to verify the study design and outcomes. The mean and standard deviation could be derived from median and range and from median and interquartile range by using the Hozo and Wan formula respectively (Gueyffier and Cucherat 2019; Hozo et al. 2005; Wan et al. 2014).

Assessment of Risk of Bias in Non-RCTs

The key difference between RCTs vs non-RCTs is the risk of bias due to confounding in the later. Assessment of the risk of bias is hence a critical step in systematic reviews of non-RCTs. The standard tools for this purpose are discussed briefly below.

(1) *The Newcastle Ottawa Scale (NOS)*

The Newcastle-Ottawa Scale (NOS) was developed by a collaboration between the University of Newcastle, Australia, and the University of Ottawa, Canada, to assess the quality of non-randomised studies (http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp).

The NOS scale contains three major domains: a selection of subjects, comparability between groups and outcome measures. The maximum score for each domain is four, two and three points, respectively. Thus, the maximum possible score for each study is 9. A total score ≤ 3 indicates low methodological quality, i.e. high risk of bias.

The NOS is a validated and an easy and convenient tool for assessing the quality of non-RCTs included in a systematic review. It can be used for cohort and case-control studies. A modified version can be used for prevalence studies. The scale has been refined based on the experience of using it in several projects. Because it gives a score between 0 and 9, it is possible to use NOS as a potential moderator in meta-regression analyses (Luchini et al. 2017; Wells et al. 2012; Veronese et al. 2016). The NOS is not without limitations. These include some of the domains that are not univocal, difficulties in adapting it to case-control and cross-sectional studies and the low agreement between two independent reviewers in scoring using NOS (Hartling et al. 2013). Training and expertise are essential for proper use of NOS (Oremus et al. 2012).

(2) *ROBINS-1 tool*

The NOS scale and the Downs-Black checklist are commonly used for assessing the risk of bias in non-RCTs. However, both include items relating to external and internal validity (Downs and Black 1998). Furthermore, lack of comprehensive manuals increases the risk of differences in interpretation by different users (Deeks et al. 2003). The ROBINS-I (“Risk Of Bias In Non-randomised Studies—of

Interventions”), is a new tool for evaluating the risk of bias in non-RCTs (Sterne et al. 2016).

Briefly, the ROBINS-1 tool considers each study as an attempt to mimic a hypothetical pragmatic RCT and covers seven distinct domains through which bias might be introduced. It uses ‘signalling questions’ to help in judging the risk of bias within each domain. The judgements within each domain carry forward to an overall risk of bias judgement across bias domains for the outcome being assessed. For details, the readers are referred to the publication by Sterne et al. (2016).

Data Synthesis

The random effects (REM) model is preferred for meta-analysis assuming heterogeneity. A categorical measure of effect size is expressed as the odds ratio (Mantel Haenszel method). Statistical heterogeneity is assessed by Chi-Squared test, I^2 statistic, and visual inspection of the forest plot (overlap of confidence intervals). The validity of REM results can be crosschecked by comparing them with the fixed-effect model (FEM) meta-analysis. Comparability of results by both models is reassuring.

While conducting meta-analysis of non-RCTs, it is important to pool adjusted and unadjusted effect size estimates separately. Pooled adjusted values must be given more importance to minimise the influence of confounders. It is important to note the type of confounders adjusted for in different studies. When synthesising results, consideration of the risk of bias in included studies is more important than the hierarchy of study design.

Publication bias: This is assessed by a funnel plot unless the number of studies is <10. Statistical tests are used if required, but their limitations need to be taken into account. It is important to note that there is no gold standard against which the funnel plot test results can be compared (Lau et al. 2006). Publication bias is not the only reason for an asymmetrical funnel plot. True heterogeneity also contributes to the small study effect (Lau et al. 2006).

Summary of findings: The data on quality of evidence, the magnitude of intervention effect, and the sum of available data on main outcomes are presented in the ‘Summary of findings table’ as per GRADE (Grading of Recommendations Assessment, Development and Evaluation) guidelines (Guyatt et al. 2013). To start with, the evidence is graded as ‘low’ given the limitations of the design of non-RCTs. It could then be upgraded based on the effect size, dose-response, and effect of all plausible confounding factors.

Important Issues in Presentation and Interpretation of Results

Understanding the properties of odds ratios (McHugh 2009; Szumilas 2010; Bland and Altman 2000; Cummings 2009; Balasubramanian et al. 2015) compared with risk ratios, the significance of unadjusted vs adjusted results, and caveats of different study designs (e.g. cohort vs case-control) is critical in presenting and interpreting the results of systematic reviews and meta-analysis of non-RCTs. It is important to note the type and number of confounders adjusted for in the included studies. Subject expertise is essential in this context. If possible, it is preferable to contact the authors of the included study for individual participant data to conduct analyses controlling for confounders. It is not unusual for the pooled effect estimates to differ based on the design of the non-RCTs. For example, pooled estimates from cohort studies have shown that red cell transfusions were associated with a lower risk of transfusion-associated necrotising enterocolitis (TA-NEC) in preterm infants. In contrast, those from case-control studies showed no association of TA-NEC with red cell transfusions (Sarooha et al. 2019).

Evidence from non-RCTs, considering their higher risk of bias, can only be used to generate hypotheses to be tested in RCTs. However, when there are no RCTs in the field of interest, non-RCTs can provide the ‘best available’ evidence for decision making. The current focus of the Cochrane collaboration on systematic reviews of non-RCTs supports this philosophy (Reeves et al. 2019).

Critical Appraisal of Systematic Reviews of Non-rCTs

AMSTAR 2 is a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions or both (Shea et al. 2017).

In summary, systematic reviews of non-RCTs are an essential part of the evidence in totality, considering RCTs may not always be available or possible for various reasons. Suppose a comprehensive literature search reveals no RCTs. In that case, a systematic review of non-RCTs is justified as long as they directly address the framed question (PICOS), and are well designed, and conducted with minimal risk of bias (Faber et al. 2016). Whether systematic reviews of non-RCTs overestimate or underestimate the effects of the intervention compared to RCTs, continues to be a controversial issue (Abrahama et al. 2010).

References

- Abraham NS, Byrne CJ, Young JM, Solomon MJ. Meta-analysis of well-designed nonrandomized comparative studies of surgical procedures is as good as randomized controlled trials. *J Clin Epidemiol*. 2010;63:238–45. <https://doi.org/10.1016/j.jclinepi.2009.04.005>.
- Balasubramanian H, Ananthan A, Rao S, Patole S. Odds ratio vs risk ratio in randomised controlled trials. *Postgrad Med*. 2015;127(4):359–67.
- Bland JM, Altman DG. The odds ratio. *BMJ*. 27 May 2000; 320: 1468.
- Szumilas M. Explaining odds ratios. *Can Acad Child Adolesc Psychiatry*. 2010; 19(3): 227–229.
- Concato J, Shah N, Horwitz RI. Randomised, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000;342(25):1887–92.
- Cummings P. The relative merits of risk ratios and odds ratios. *Arch Pediatr Adolesc Med*. 2009;163(5):438–45.
- Deeks JJ, Dinnes J, D'Amico R, et al. International Stroke Trial Collaborative Group European Carotid Surgery Trial Collaborative Group. Evaluating non-randomised intervention studies. *Health Technol Assess*. 2003;7:iii–x, 1–173.
- Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health*. 1998;52:377–84.
- Faber T, Ravaut P, Riveros C, Perrodeau E, Dechartres A. Meta-analyses including non-randomized studies of therapeutic interventions: a methodological review. *BMC Med Res Methodol*. 2016;16:35. <https://doi.org/10.1186/s12874-016-0136-0>.
- Faraoni D, Schaefer ST. Randomised controlled trials vs. observational studies: why not just live together? *BMC Anesthesiol*. 2016 Oct 21;16(1):102.
- Gershon AS, Jafarzadeh SR, Wilson KC, Walkey A. Clinical knowledge from observational studies: everything you wanted to know but were afraid to ask. *Am J Respir Crit Care Med*. 2018; 198 (7):859–867.
- Gerstein HC, McMurray J, Holman RR. Real-world studies no substitute for RCTs in establishing efficacy. *Lancet*. 2019;393:210–1.
- Gilmartin Thomas JFM, Liew D. Observational studies and their utility for practice. *Aust Prescr*. 2018;41:82–5.
- Gueyffier F, Cucherat M. The limitations of observation studies for decision making regarding drugs efficacy and safety. *Therapie*. 2019;74:181–5.
- Guyatt GH, Oxman AD, Santesso N, et al. GRADE guidelines: 12 Preparing summary of findings tables—binary outcomes. *J Clin Epidemiol* 2013;66:158–172.
- Hartling L, Milne A, Hamm MP, Vandermeer B, Ansari M, Tsertsvadze A, Dryden DM. Testing the Newcastle Ottawa Scale showed low reliability between individual reviewers. *J Clin Epidemiol*. 2013;66:982–93.
- Heikinheimo O, Bitzer J, Rodríguez LG. Real-world research and the role of observational data in the field of gynaecology—a practical review. *Eur J Contracept Reprod Health Care*. 2017;22 (4):250–9.
- Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol*. 2005;5:13.
- Jepsen P, Johnsen SP, Gillman MW, Sorensen HT. Interpretation of observational studies. *Heart*. 2004;90(8):956–60.
- Joffe MM, Rosenbaum PR. Invited commentary: propensity scores. *Am J Epidemiol*. 1999;150 (4):327–33.
- Lau J, Ioannidis JP, Terrin N, et al. The case of the misleading funnel plot. *BMJ*. 2006;333:597–600.
- Lefebvre C, Manheimer E, Glanville J. Searching for studies. *Cochrane handbook for systematic reviews of interventions*. New York: Wiley 2008:95–150.
- Lefebvre C, Glanville J, Wieland LS, et al. Methodological developments in searching for studies for systematic reviews: past, present and future? *Syst Rev*. 2013;2:78.

- Ligthelm RJ, Borzi V, Gumprecht J, Kawamori R, Wenying Y, Valensi P. Importance of observational studies in clinical practice. *Clin Ther.* 2007;29 Spec No:1284–92.
- Luchini C, Stubbs B, Solmi M, Veronese N. Assessing the quality of studies in meta-analyses: Advantages and limitations of the Newcastle Ottawa Scale. *World J Meta-Anal.* Aug 26, 2017; 5(4): 80–84. Published online Aug 26, 2017. <https://doi.org/10.13105/wjma.v5.i4.80>.
- Mann CJ. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emerg Med J.* 2003;20(1):54–60.
- Mariani AW, Pego-Fernandes PM. Observational studies: why are they so important? *Sao Paulo Med J.* 2014;132(1):01–02 https://www.scielo.br/scielo.php?script=sci_arttext&pid=S1516-31802014000100001&lng=en&tlng=en. Accessed 10 Aug 2020.
- McHugh ML. The odds ratio: calculation, usage and interpretation. *Biochem Med.* 2009;19 (2):120–126.
- Norris S, Atkins D, Bruening W, et al. Selecting observational studies for comparing medical interventions. In: Agency for Healthcare Research and Quality. Methods Guide for Comparative Effectiveness Reviews [posted June 2010]. Rockville, MD. http://www.effectivehealthcare.ahrq.gov/ehc/products/196/454/MethodsGuideNorris_06042010.pdf. Accessed 11 Aug 2020.
- Oremus M, Oremus C, Hall GB, McKinnon MC; ECT & Cognition Systematic Review Team. Inter-rater and test-retest reliability of quality assessments by novice student raters using the Jadad and Newcastle-Ottawa Scales. *BMJ Open.* 2012;2:e001368.
- Reeves BC, Deeks JJ, Higgins JPT, et al. Chapter 24: Including non-randomised studies on intervention effects. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.0 (updated July 2019). Cochrane, 2019. www.training.cochrane.org/handbook. Accessed 10th Aug 2020.
- Saroha V, Josephson CD, Patel RM. Epidemiology of necrotising enterocolitis: New considerations regarding the influence of red blood cell transfusions and anemia. *Clin Perinatol.* 2019;46(1):101–17. <https://doi.org/10.1016/j.clp.2018.09.006>.
- Schillaci G, Battista F, Pucci G. Are observational studies more informative than randomised controlled trials in hypertension? *ConSide of the Argument.* Hypertension. 2013;62:470–6.
- Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ.* 2017;358: <https://doi.org/10.1136/bmj.j4008> (Published 21/9/2017).
- Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ.* 2016;355:i4919.
- Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA.* 2000;283:2008–12.
- Veronese N, Carraro S, Bano G, Trevisan C, Solmi M, Luchini C, Manzano E, Caccialanza R, Sergi G, Nicetto D. Hyperuricemia protects against low bone mineral density, osteoporosis and fractures: a systematic review and meta-analysis. *Eur J Clin Invest.* 2016;46:920–30.
- Wan X, Wang W, Liu J, et al. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol.* 2014;14:135.
- Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of non-randomised studies in meta-analyses, 2012. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.