PRECIS-2 for retrospective assessment of RCTs in systematic reviews: some thoughts on intention, dichotomization and applicability of RCTs.

Merrick Zwarenstein,
Centre for Studies in Family Medicine, Department of Family Medicine, Western University, London, Ontario
IC/ES, London, Ontario.
merrick.zwarenstein@ices.on.ca

Kevin Thorpe,
Applied Health Research Centre, Li Ka Shing Knowledge Institute of St Michaels Hospital, Toronto, Ontario
Dalla Lana School of Public Health, University of Toronto, Toronto Ontario
kevin.thorpe@utoronto.ca

Shaun Treweek,
Health Services Research Unit, University of Aberdeen, Aberdeen, UK
streweek@mac.com

Kirsty Loudon,
Freelance researcher,
Kirsty.loudon1@btinternet.com

PRECIS-2 in systematic reviews: some thoughts on retrospective assessment of randomized trials for intention and applicability.

Introduction

The Journal of Clinical Epidemiology is the main forum for discussions on prospectively matching pragmatic or explanatory intent with the design of single randomized controlled trial (RCTs). With this paper by Dal-Ré et al[1] it has also become a forum for discussing the PRECIS-2 tool[2] for retrospectively assessing multiple RCTs within reviews. PRECIS derives from the acronym "PRagmatic Explanatory Continuum Indicator Summary", with the word pronounced as "pray-see" and meaning "summary". Although Dal-Ré uses examples of sub-discipline reviews, we discuss also systematic reviews of intervention effectiveness, into which PRECIS-2's measure of design features and usefulness for choosing between interventions could be incorporated. This would complement the Cochrane Risk of Bias score[3], which permits users to consider the quality of trials in their decision.

Dal-Ré proposes that any one of 3 features (a placebo control group, masking/blinding of participants or providers, or any RCT conducted only in a single centre) should automatically categorize the entire trial and its findings as maximally explanatory, without assessing the degree of pragmatism of the nine domains of PRECIS-2. Dal-Ré also challenges the use of the main peer reviewed trial report as the main documentary source on which to base retrospective assessment of an RCT and raises questions about the potential for bias if the design features of a trial appear to have changed between information sources. Finally, he proposes that, where information is missing or unclear that domain should be scored as a blank. We respond to these constructive suggestions in order.

An RCT with placebo control or blinding cannot support decision making

We consider these two characteristics together, as placebo control is functionally connected with blinding. In combination these eliminate the placebo effect, the effect of subjective belief in an intervention on outcome. Even though this effect is generally small[4], placebo control with blinding has become an RCT design default even when not needed, e.g., in RCTs with objective outcomes. Dal-Ré proposes that any RCT that uses a placebo comparator or blinds the participants (patients and/or providers) in a trial departs so far from the usual care in that setting that the RCT as a whole should automatically be classified as maximally explanatory, without further assessment of PRECIS-2 domains. We see this as a question of the appropriateness of dichotomization as pragmatic or explanatory for the intentions underlying a trial, for the trial as a whole, and for individual design features.

Schwartz and Lellouch describe the intention of the designers for their trial as a dichotomy, either predominantly pragmatic or predominantly explanatory; and individual features as potentially a mixture of more pragmatic or more explanatory; but they almost always avoid dichotomizing any trial as a whole as pragmatic or explanatory. PRECIS-2 authors agree, focusing on individual design features, adding the nuance that each feature is on a continuum between explanatory and pragmatic, and steering away from assessing the trial as a whole.

For Schwartz and Lellouch the starting point for pragmatism is the intention to design a trial whose findings are useful for decision-making, followed by selection of design

characteristics to match that intention. They describe placebo control and blinding as a design choice typical of an explanatory rather than a pragmatic intention, but they also emphasize a multi-factorial approach to the match between intention and design, in which each of several characteristics may align to a greater or lesser extent with the intention of the RCT. While we agree that placebo treatment with blinding is seldom part of real-world care, there are situations in which a trial could include a placebo control with blinding, while still providing information that directly informs a decision.

Consider a two arm, placebo-controlled trial of a new treatment (with unknown risks, or expensive), aimed at treating a non-life-threatening condition where the outcomes are participant reported and amenable to subjective influence. Or consider a comparative effectiveness study in which two or more widely accepted treatments are compared with each other, with placebos used to disguise the treatment arm from participants to ensure that the choice is based not on subjective belief but on biological activity. Decision makers balance benefits, costs and harms when the choice between interventions is close. Eliminating patient or provider subjectivity as an explanation for effectiveness by use of placebos and blinding might provide information that reassures the decisionmaker that the apparent superiority of one intervention is not due to subjectivity alone and is thus worth its costs and risks. Trials providing such insights fit Schwartz and Lellouch's definition of pragmatic intention.

Review authors should not therefore prematurely categorize any included trial as explanatory simply because they believe a single design feature overrides others. Doing so supposes that such trials are never useful (see above), that there is strong empirical evidence on the relative weight of each design feature in regard to usefulness for decision makers, and that it is valid and useful to dichotomize a trial as a whole. PRECIS-2 authors are unaware of such empirical evidence, and we and Schwartz and Lellouch would avoid labelling trials in their entirety as dichotomously pragmatic or explanatory (preferring that decision makers consider the individual design features in all trials in the review). Every included trial should therefore be described without prejudice in a review, so that the decisionmakers can assess the usefulness to their own choice of each trial according to its similarity to their situation on individual design features (PRECIS-2 domains)

Although we do not agree that placebo comparators and blinding overrides other design features, we are persuaded by Dal-Re's argument that this information should be included in PRECIS-2, which currently addresses placebo control and blinding only by urging that a trial with pragmatic intentions be designed with a usual care comparator. This does not work for retrospective use of PRECIS-2 in reviews. Since we prefer to keep a single version of the PRECIS-2 tool that can be used for both design and for retrospective assessment of RCTs for reviews, we propose that' Dal-Ré's concern regarding the potential distorting effect of placebo control and blinding be addressed by re-introducing comparator choice as a separate new 10th domain in a PRECIS-2 update.

In the precursor to PRECIS-2[5] two overlapping domains addressed the comparator: comparator flexibility and comparator provider expertise. We propose a single domain where the pragmatic end refers to control groups whose participants' care replicates the mix which constitutes usual care in the setting in which the trial was conducted, while the

explanatory end describes trials with placebo control groups and blinding of participant and intervention providers. An active treatment that is usual care, not necessarily in the trial setting, could be scored as highly pragmatic on this domain. Blinding of an unobtrusive independent assessor is also a pragmatic design choice. In a trial with multiple arms, or multiple controls some of which are placebos, we suggest that the PRECIS-2 tool be used once for each head-to-head comparison between arms, rather than just once for each trial[6].

A single centre trial cannot contribute to decision making

Dal-Ré proposes that a trial conducted only in a single centre should automatically be characterized as fully explanatory. We see this as a question of external validity- how do we apply the results of trials with explanatory and pragmatic features to situations outside of the trial itself?

Schwartz and Lellouchs' seminal insight was that RCTs could have one of two intentions or purposes, pragmatic or explanatory, and that most trial designers defaulted unthinkingly to the latter[7]. Pragmatic intentions should lead to design of a trial whose findings directly inform decisionmakers in the setting of that trial of the average benefits and harms of the interventions, and also directly inform external decisionmakers elsewhere, provided that the trial sites are a statistically representative sample of the larger universe for which those decision makers are responsible. Beyond these exceptions external decisionmakers need to judge for themselves whether their situation is similar enough on major contextual features to that of the trial that the same outcomes are likely. By completing a PRECIS-2 appraisal of a trial with pragmatic intentions, external decisionmakers can see that few or no design restrictions (on domains such as participant inclusion, flexibility and resources for intervention delivery, outcomes and intensity of measurement) were added to complicate their judgement. The judgement of similarity and thus applicability is then straightforward, although never guaranteed, given unknown or unmeasured differences between the trial and their own context.

By contrast, explanatory intentions lead to design of a trial whose findings deepen scientific understanding of a hypothesized mechanism of action of an intervention. Because of the design restrictions imposed to highlight this mechanism, the situation of an explanatory trial is not similar to any real-world setting, and so the trial findings cannot be directly applied anywhere, not even in the setting of the trial itself (once the trial is over and usual care returns). Instead, application of findings from a trial designed with explanatory features must be based on the argument that the mechanism demonstrated to operate in the trial is dominant over any contextual differences in participants, delivery, outcome measures or other factors in the situation to which its findings are being applied. This judgement of applicability must be based upon many counterfactual assumptions and is therefore less direct and thus less likely to be reliable than for a trial with more pragmatic features.

We agree that evidence from a trial with pragmatic design choices and multiple centres suggests wide applicability of its findings and is reassuring for an external decisionmaker. But there will also be decisionmakers for whom the findings of a single centre trial are useful, either because they are making decisions for that centre itself post-trial, or because they are confident that their own situation is very similar to that single centre. With this potential overriding feature, as with placebo and blinding, we feel that prejudging some

trials in a systematic review as fully explanatory on the basis of a single overriding characteristic would eliminate from consideration those that might, aside from their number of sites, be pragmatic in intention and in other domains. We also want to point out that consideration of the number of sites in a trial is already assessed under Domain 3 in the PRECIS-2 tool.

PRECIS-2 cannot be assessed retrospectively using only the main published RCT report
Dal-Ré points out that the main trial report may often leave out important details and argues that this precludes reliable assessment of the items in the PRECIS-2 tool. We see this as an issue of quality control on trial reports.

There is a well-recognized, repeatedly updated, and evidence based template for the main trial report: the CONSORT statement, with multiple subtypes for specific kinds of RCTs, including those defined by structure, such as cluster randomized or stepped wedge trials, and even for specific intentions, such as our CONSORT extension for trials with pragmatic intent. Not only are these templates widely accepted by trial researchers, they are required by many journals at submission. This pressure for the published trial report to be highly standardized may be improving quality of reporting[8]. We expect that with further updates, many of the information gaps in published reports of RCTs will shrink. This contrasts with trial registration: although required by journals, only a bare minimum of information is entered into the registration website, much less than in the final report; and although SPIRIT, the trial protocol template is standardized there is no requirement that protocols use it. We agree with Dal-Ré that the trial report is less than perfect but in balancing these considerations we propose that retrospective assessments using PRECIS-2 for systematic reviews continue to rely primarily on the main published peer reviewed report of each RCT. Trial designers could make the task of reviewers and retrospective assessors easier by using the CONSORT extension for Pragmatic trials[9] to ensure that all the key information specific to RCTs with pragmatic intent is included in their trial report; and, as Dal-Rè has suggested, by also reporting their own PRECIS-2 assessment in that report.

Focusing on the published main RCT report avoids the need for dealing with changes to trial descriptions that may have arisen between earlier documents (trial registration record, trial protocol) and the peer reviewed published final paper. While this may incur some risk of using erroneous or falsified information, it is not practical given the length of the usual protocol (tens to hundreds of pages) for the researcher conducting a systematic review to act as a forensic reader and compare every document describing every trial, even less so for decisionmakers including clinicians, health system managers, or policy makers. Since use of these other documents is not mandatory for high quality Cochrane reviews it seems out of place to insist on it for the PRECIS-2 assessment. Cochrane reviews already require massive effort and if PRECIS-2 is to be integrated into Cochrane systematic reviews, which is our hope, we need to take care not to add work without evidence of value.

Dealing with missing information on PRECIS-2 domains
On occasion, the published RCT report fails to provide enough information to assess (retrospectively) the score for one or more domains. There are a few situations in which a domain may be justifiably excluded from a trial and PRECIS-2 recommends it be left blank. If one of the domains is irrelevant, e.g. efforts to maintain adherence on the part of the

patient where the intervention is defined as receipt of a single surgical procedure, it should be left blank. Although rare, if a trial is being designed to evaluate an intervention which directly aims to change an outcome that is also used to score a domain of the PRECIS-2 tool (for example, adherence[10]) that domain should be left blank. But Dal-Ré has identified a contradiction between our recommendation for scoring PRECIS-2 domains with inadequate information during design (leave blank and explain reason in table or text), and when retrospectively assessing trials in a review (score as a middle value: 3). Dal-Ré points out that this latter approach biases the domain score towards 3 and recommends instead that missing information from a trial in a review should be treated as it would be in a trial during design, and left blank. We agree and will include this in the next update of the PRECIS-2 tool. The denominator for the total PRECIS-2 score for such a trial should match the number of domains for which data is available. There would be fewer blank domains if trial designers using PRECIS-2 also used the CONSORT statement for reporting their completed RCT.

Conclusion

We would like to thank Dal-Ré and others in this community for their longstanding and constructive engagement with the ideas of Schwartz and Lellouch. The value of randomization to trials was quickly obvious, but it has taken longer for their insight into pragmatic and explanatory intentions to be absorbed. We look forward to the realization of their vision: "…..(T)he distinction between the two kinds of research, the one aiming to extend our field of knowledge, the other aiming at rational decision making, concerns an area far broader than that of clinical trial. …….(G)eneralizing the distinction to the broadest fields of research cannot but lead to fruitful consequences"[11].

---

[1] The design can limit PRECIS-2 retrospective assessment of the clinical trial explanatory/pragmatic features Dal-Ré R, de Boerb A, James SK. Journal of Clinical Epidemiology, 2020.

[2] Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ 2015;350:h2147.

[3] Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, Cates CJ, Cheng H-Y, Corbett MS, Eldridge SM, Hernán MA, Hopewell S, Hróbjartsson A, Junqueira DR, Jüni P, Kirkham JJ, Lasserson T, Li T, McAleenan A, Reeves BC, Shepperd S, Shrier I, Stewart LA, Tilling K, White IR, Whiting PF, Higgins JPT. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019; 366: l4898.

[4] Hróbjartsson A, Gøtzsche PC. Placebo interventions for all clinical conditions. Cochrane Database of Systematic Reviews 2010, Issue 1. Art. No.: CD003974. DOI: 10.1002/14651858.CD003974.pub3.

[5] Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. J Clin Epidemiol. 2009;62(5):464-475. doi:10.1016/j.jclinepi.2008.12.011

6 Lipman, P.D., Loudon, K., Dluzak, L. et al. Framing the conversation: use of PRECIS-2 ratings to advance understanding of pragmatic trial design domains. Trials 18, 532 (2017). https://doi.org/10.1186/s13063-017-2267-y

[7] Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. J Chronic Dis. 1967;20:637–48.

[8] Turner L, Shamseer L, Altman DG, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. Cochrane Database Syst Rev. 2012;11:MR000030. Published 2012 Nov 14.

[9] Zwarenstein M, Treweek S, Gagnier JJ, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ. 2008;337:a2390. Published 2008 Nov 11. doi:10.1136/bmj.a2390

[10] Zwarenstein M, Schoeman JH, Vundule C, Lombard CJ, Tatley M. Randomised controlled trial of self-supervised and directly observed treatment of tuberculosis. Lancet. 1998;352(9137):1340-1343. doi:10.1016/S0140-6736(98)04022-7

[11] Schwartz D, Flamant R and Lellouch J. Transl: Healy MJR: Clinical Trials 1980. p 253-4 Academic Press, London