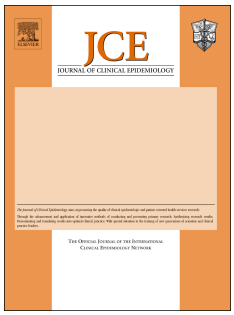
Accepted refereed manuscript of:

Zwarenstein M, Treweek S & Loudon K (2017) PRECIS-2 helps researchers design more applicable RCTs while CONSORT Extension for Pragmatic Trials helps knowledge users decide whether to apply them. *Journal of Clinical Epidemiology*, 84, pp. 27-29.

DOI: 10.1016/j.jclinepi.2016.10.010

Accepted Manuscript



Understanding pragmatism and PRECIS-2

Merrick Zwarenstein, Shaun Treweek, Kirsty Loudon

PII: S0895-4356(16)30410-3

DOI: 10.1016/j.jclinepi.2016.10.010

Reference: JCE 9315

To appear in: *Journal of Clinical Epidemiology*

Received Date: 4 September 2016

Accepted Date: 8 October 2016

Please cite this article as: Zwarenstein M, Treweek S, Loudon K, Understanding pragmatism and PRECIS-2, *Journal of Clinical Epidemiology* (2017), doi: 10.1016/j.jclinepi.2016.10.010.

Understanding pragmatism and PRECIS-2

Merrick Zwarenstein
Director, Centre for Studies in Family Medicine,
Department of Family Medicine
Schulich School of Medicine & Dentistry
Western University
1151 Richmond St
London, ON N6A 3K7
Canada

Shaun Treweek
Health Services Research Unit, University of Aberdeen, UK

Kirsty Loudon, Nursing, Midwifery, Allied Health Professions Research Unit., University of Stirling, UK

1872 words

We appreciate the efforts of the editors of the Journal of Clinical Epidemiology to provide a forum for discussion of pragmatism[i] in randomized trials (RCTs), with many items in recent years, including three in this issue by Dekkers et al [ii], Zuidegeest et al. [iii] and Riddle [iv]. We are encouraged that these three authors agree on the importance of pragmatism and the relevance of PRECIS-2 to RCT design[v], our tool to promote better matching of the choices made during design with the intended use of that trial's results. We thank them for their insightful comments relating to the varied meanings of generalizability, and the relationship between pragmatic and explanatory characteristics and internal validity and applicability (which we use as a synonym for generalizability) of trial results, to which we respond below.

Dekkers et al conclude their commentary as follows: "Unfortunately, the PRECIS authors suppose that a pragmatic trial is widely generalizable, to different contexts and settings…". We do not suppose this. In fact Dekkers et al correctly describe our approach in an earlier sentence in their paper in which they refer to PRECIS-2 as having the 'objective to match the trial design ''to how the trial results are intended to be used''. Our conception of PRECIS-2 is that it is to be used by trialists to design a trial whose results are applicable to a context in which they, *the trialists,* are intending the results to be used. At the design stage, trialists carefully consider the population and health care delivery context required for their intervention to be successful, use this to formulate a realistic view of the applicability of their trial results, and then match the trial design to that context. The full quote from the PRECIS-2 paper is as follows: "The PRECIS-2 tool focuses on trial design choices which determine the applicability of a trial. Applicability (the ability for a trial result to be applied or used in a particular situation) is the outcome of these choices, which affect the ease with which the trial results can be applied to and by the usual community of users of the intervention in the settings in which the trial designers envisioned it

being used". This misunderstanding by Dekkers et al and others is a clear signal that we need to be more explicit about the process by which designers formulate their intentions for applicability in our next update of PRECIS.

Dekkers et al end their paper by stating that "one may question whether ……. trialists are in a position to judge such a broad applicability of their own trial". We do believe that a pragmatic trial, designed to be applicable to a particular population and health system context, and using recruitment, follow up, analysis and other design choices that most closely mimic care delivery in that context, may well be applicable to other *similar* contexts that lie well outside of the trial designers' experiences or intentions. But this judgement of similarity, and thus of applicability is really a decision about local implementation, and is not the responsibility of the trial design team, but of the *reader* who may be a decision maker and potential user of the evidence provided in the published RCT. We long ago designed a different tool intended to ease the task of such readers in making decisions using RCT publications. The CONSORT extension for Pragmatic Trials[vi] is intended for use when the results of a trial are being written up for publication. It provides guidance to the writers of that publication, presumably part of the trial design team, so that they include in their trial publication all the information on the context of their trial needed by future readers of that paper to judge applicability of the results to their own context, where they may be considering implementing the intervention tested in that randomized trial.

In summary, the issue of applicability is dealt with in two stages, by different actors. In the first stage, trial designers use the PRECIS-2 tool to match their trial design to a context in which they, the trial team, deem the intervention would be usable and the RCT results applicable. At the second stage, health system decision makers use the information in the published RCT to contribute to the decision on implementation of the tested intervention in their own context. This also answers the questions raised by Dekkers et al. as to whether a trial can be equally applicable in Finland and China, in primary care and in intensive care: the trial designers are unlikely to be able to judge applicability between two such different countries, but we would speculate that intensive care units may be more similar to each other across countries than are their primary care systems. Mainly, we suggest that these are second stage judgements, to be made by local decision makers in any setting considering implementation of the tested intervention. In other words, separately by Finnish and Chinese decision makers as they consider implementing the same intervention in their own contexts. Most often this requires that the decision makers return to the published RCT of that intervention, as well as having a nuanced understanding of any existing systematic reviews of RCTs of that intervention, which might be aggregating outcomes across very different contexts.

We agree with Zuidgeest et al that designing a pragmatic trial would be difficult if every imaginable characteristic of the context to which the trial is generalized had to be explicitly matched, and that it is acceptable to ignore matching on characteristics that are known *not* to affect outcome. Unfortunately, we seldom know the full pathway of action for drugs, much less that for complex interventions operating through multiple pathways, and so it is very hard to know which potential matching characteristics can be ignored. For this reason we implicitly prioritized a number of features in PRECIS-2 and focussed on matching aspects typically under the control of the designers of the trial themselves to the context in which they deem the trial applicable: these nine characteristics form the domains of PRECIS-2: eligibility criteria, recruitment processes and intensity of follow up of participants, the rigidity with which delivery and participant adherence are enforced, the approach to primary outcome and analysis, and two, setting and organization, intended to capture the local peculiarities of the health systems and contexts in which the designers consider that trial results will be applicable. We do not prescribe detailed criteria for these last two domains because at this stage we do not know how specific attributes

influence generalizability. We look forward to the Get Real Consortium work, by Zuidegeest et al, on the effects on trial conduct, ethics and feasibility of different design choices, as it may provide empirically tested and robust guidance to trial designers on different options for balancing practical and pragmatic aspects of their proposed trial. But PRECIS-2 users, such as trial designers, funders and readers of randomized trials should keep in mind that the act of generalization is likely to remain an act of subjective judgement, not statistical extrapolation.

Zuidgeest et al correctly point out that the PRECIS-2 paper says little about the comparator arm, and excludes it from the PRECIS wheel even though choice of comparator may dramatically influence the estimate of effect size. We agree that the choice of comparator is difficult and critical and yet we say only: “In pragmatic trials the comparator is usual care. In explanatory trials it may not be. In PRECIS-2 the domains are based on the assumption that the trial is two armed, one of which is usual care with no changes. If usual care is not the comparator, or there are multiple intervention arms that are very different from each other, then the arms will need to be scored separately”. We suggest that the definition of usual care (or the chosen comparator) is the responsibility of the trial design team, as is the design of intervention arm(s), and we don't presume to direct the trial design team on what they should select as a comparator. We find it difficult to suggest more than this without narrowing the value of PRECIS-2 by recommending one or other “standard” for usual care which may even slant guidance towards particular settings. Detailed description of the comparator arm is crucial to the second stage judgement of applicability mentioned above and for this reason our CONSORT extension for pragmatic trials specifies that the comparator arm(s) should be described in as much detail as the intervention arm(s).

Riddle points out that there are times when biological mechanisms, rather than “real world” impact of interventions must be studied, and that explanatory approaches to trial design are then completely appropriate. We agree. We also agree with Riddle that trial designers need to keep a careful eye on trial quality, in relation to internal validity. These issues lie outside the scope of PRECIS-2 but we concur that a pragmatic trial conducted poorly, in ways that undermine internal validity, is of little use for decision support. However, we do not agree with Riddle that trial design choices such as definition of usual care, and treatment standardization have any impact on internal validity; indeed, we would argue that these particular items impact only on external validity/applicability/generalizability. Trialists need a nuanced understanding of which issues in design are linked to internal validity (bias), and which to external validity (generalizability, applicability), in order to design their trials appropriately.

Zuidgeest et al correctly point out that PRECIS-2 aims to help all trial designers match their design to the intended purpose of their trial, but its’ descriptions of the attributes that promote an explanatory approach are less detailed than those which promote a pragmatic approach. They suggest that this is not helpful since the characteristics that promote clear testing of an hypothesis about mechanism are unlikely to be the simple inverse of those which promote applicability. This is a valid criticism, and at our next update of PRECIS we will consider how best to support trialists whose intentions are mechanism-oriented and explanatory. We will also consider this issue as we update the CONSORT statement on pragmatic trials, where we encourage trialists to match the type of conclusions they draw to the capacities of their trial design to support those conclusions.

It is, however, worth mentioning that we see many RCTs published with explanatory designs, and yet they conclude with a treatment recommendation, which suggests that hypothesis testing or mechanism exploration may not have been their main goal. Rather, the presence of a treatment recommendation suggests a pragmatic (decision support) purpose for the trial and that the

choice of explanatory design elements represents a mismatch between the trialists' applicability intentions and their design choices. PRECIS-2 is intended to help all trialists match their design choices to their intentions, but the most valuable contribution of PRECIS-2 to health systems would be to prevent this longstanding problem: trials whose intent is pragmatic but whose design choices are skewed towards the explanatory end of the spectrum. We quote Schwartz and Lellouch: "Most trials done hitherto have adopted the explanatory approach without question; the pragmatic approach would often have been more justifiable".

[i] Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. Republished in: J Clin Epidemiol 2009;62:499e505. Original publication: J. chron. Dis. 1967, Vol. 20, pp. 637-648.

[ii] Zuidgeest, MG., Goetz, I., Grobbee, DE. PRECIS-2 in perspective: what's next for pragmatic trials? Jnl Clin Epidemiol 2016 (This issue?)

[iii] Riddle, DL., Consequences of randomized clinical trial design decisions need to be clarified. Jnl Clin Epidemiol 2016 (This issue?)

[iv] Dekkers, OM., Bossuytd, PM., Vandenbroucke JP. How trial results are intended to be used: is PRECIS-2 a step forward? Jnl Clin Epidemiol2016 (This issue?)

[v] 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.

[vi] Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, Gent M, Oxman AD, Moher D for the CONSORT and Pragmatic Trials in Healthcare (Practihc) groups. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ 2008a2390 doi: 10.1136/bmj.a2390

Highlights

RCTs have been the mainstay of evaluation of health and medical interventions for over half a century, but doubts on their applicability to real world contexts has been an increasing topic of discussion in recent years. We comment here on the understandings of PRECIS-2 described in three papers in this issue (Zuidegeest, Dekkers, and Riddle) focussing on the issue of who judges applicability and when. We point out the focus of PRECIS-2 on designing trials to match the intentions of the designers, which might be either applicability to a specific context that the designers have in mind, or testing a hypothesis related to a mechanism of action. PRECIS-2 is suitable for this purpose, but a different tool, the CONSORT statement extension for pragmatic trials is suitable to support the proper description of intervention and context by the authors of the trial so that the information in it is of maximum usefulness to potential users or implementers of the tested intervention in settings beyond those imagined by the trials designers.