

**TITLE:** The importance of decision intent within descriptions of pragmatic trials

**AUTHORS:** Stuart G. Nicholls, Merrick Zwarenstein, Spencer, P. Hey, Bruno Giraudeau, Marion K. Campbell, Monica Taljaard

**CORRESPONDING AUTHOR:** Stuart G. Nicholls, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Civic Campus, 1053 Carling Ave, Ottawa, ON K1Y 4E9  
T: 613-798-5555 ext.19640; Email: [snicholls@ohri.ca](mailto:snicholls@ohri.ca)

**MZ:** Centre for Studies in Family Medicine, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada [merrick.zwarenstein@ices.on.ca](mailto:merrick.zwarenstein@ices.on.ca)

**SPH:** Center for Bioethics, Harvard Medical School, Boston MA, USA [spencer\\_hey@hms.harvard.edu](mailto:spencer_hey@hms.harvard.edu)

**BG :** Université de Tours, Université de Nantes, INSERM, SPHERE U1246, Tours, France; INSERM CIC1415, CHRU de Tours, Tours, France. [bruno.giraudeau@univ-tours.fr](mailto:bruno.giraudeau@univ-tours.fr)

**MC :** Health Services Research Unit, University of Aberdeen, Health Sciences Building, Foresterhill, Aberdeen, UK, [m.k.campbell@abdn.ac.uk](mailto:m.k.campbell@abdn.ac.uk)

**MT:** Ottawa Hospital Research Institute (OHRI), Ottawa, Ontario, Canada; School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada [mtaljaard@ohri.ca](mailto:mtaljaard@ohri.ca)

**WORD COUNT:** 3530 words (plus tables and references)

**Abstract (current: 212 words, MAX 300 words)**

It is now more than 50 years since the concepts of explanatory and pragmatic attitudes towards trials were first discussed by Schwartz and Lellouch in their influential 1967 paper. Since then there has been increasing focus on design aspects that may be consistent with more pragmatic attitudes within clinical trials, and a number of tools developed to assist investigators prospectively think about their trial design. Researchers have subsequently expressed interest in using these tools retrospectively to characterise trials as pragmatic or explanatory. We suggest that recent attempts to retrospectively dichotomise trials solely on the basis of quantitative scoring of trial design features are flawed. Instead, we argue that there is a need to consider both the intent and design when assessing the degree of pragmatism within a trial. The practical implication of our suggestion for trial reporting is that investigators should explicitly state the intent of the trial through a clear articulation of the decision that they hope will be informed by the trial results. This should be coupled with a completed PRECIS-2 assessment (or similar) with an explanation of study design choices, in order to appropriately assess whether the study design is consistent with the study intent. We believe this will assist reviewers and knowledge users in making assessments of trials.

**Keywords**

Randomized controlled trial; classification; quantitative assessment, retrospective

## **Introduction and background**

It is now more than 50 years since the concepts of explanatory and pragmatic attitudes towards trials were first espoused by Schwartz and Lellouch (1967) in their influential paper on attitudes toward therapeutic trial design. These authors argued that approaches to study design, the question of interest, and the type of knowledge sought from the study are key elements to the “attitude” of a trial. As paraphrased by Zwarenstein and Treweek, pragmatic attitudes are those that seek to inform real world decisions regarding alternative treatments options while explanatory attitudes seek understand the mechanism of action of an intervention (Zwarenstein & Treweek, 2009)

Interest in pragmatic randomized controlled trials (pragmatic RCTs) has increased substantially in recent years, and notably so since the turn of the century (Chalkidou et al., 2012; Patsopoulos, 2011). The increased interest is likely due to the need by decision makers at policy and clinical levels for more relevant and applicable research, and the needs of research funders to demonstrate the contribution of research tax dollars to health improvements. Further, there is concern that many trials have failed to predict the actual effectiveness of an intervention in later clinical practice, and diluted effects have been observed when interventions have been rolled out to a broader clinical population. In part, this has been attributed to the disconnect between their intent and their design (Ford & Norrie, 2016). Others have noted that explanatory RCTs often exclude individuals who would likely receive a study intervention in practice, leading to a lack of good quality evidence to inform many treatment decisions for these populations (Roland & Torgerson, 1998). While some commentators believe that particular design features, such as blinding or placebo controls, may be sufficient to rule out a trial as pragmatic, others have argued that such features should not be defining characteristics of a pragmatic trial (Dal-Ré, 2019; Dal-Ré et al., 2018; Sedgwick, 2014; Kevin E. Thorpe et al., 2010; K. E. Thorpe et al., 2009; Zwarenstein et al., 2008). As such, whether certain features would exclude a trial from being pragmatic remains an area of debate.

As interest in pragmatic RCTs has increased, a number of tools (Koppelaar et al., 2011; Loudon et al., 2015; K. E. Thorpe et al., 2009) have been developed to assist investigators evaluate the appropriateness of design decisions when thinking about their trial design. Notably, these tools have increasingly involved quantitative elements that allow trials to be scored across domains. While these tools have been developed for prospective use, researchers have also expressed interest in retrospectively applying these tools to trial reports in order to evaluate and classify trials as pragmatic or not (Sajobi et al., 2018; Yoong et al., 2014). Thus, there has been increasing focus on the quantification of design features and attempts to categorise trials based purely on these design features.

In the present analysis we highlight, and critique, two separate but related aspects of this focus. First, we argue that it is imperative that the trial attitude (or what we shall call the trial intent) is integrated with consideration of the design features. This contention is based on our belief that the design of a trial should flow from its intent (and the decision to which the results of the trial are intended to be applicable) and thus a focus purely on retrospective quantitative evaluation, and classification based on this score, ignores the trial intent and potential applicability of the findings. Second, we note that there remain a number of conceptual and practical impediments to categorisation of trials based on retrospective scoring of trial design characteristics, not least the lack of consensus regarding the placement of thresholds (if at all), weighting of different trial characteristics, and reported variation in the scoring of individual characteristics.

The structure of the manuscript is as follows: We first review tools developed for investigators to evaluate trials in relation to their intent. We then critique how use of these tools has evolved. We conclude by arguing that trial evaluations cannot be judged solely on an abstract retrospective review of design features but requires both an understanding of the intent of the trial and how the design relates to this intent and the decision for which the trial results were intended to be applicable.

## **Specifying the domains of pragmatic trials: the development of frameworks and tools for prospective assessment**

Since Schwartz and Lellouch (Schwartz & Lellouch, 1967) first promoted the idea of pragmatic and explanatory attitudes toward trials, there has been increasing attempts to formalise and quantify trial design features that are consistent with a pragmatic intent (see Supplementary Material S1 for examples and Loudon et al., 2013a; Pawson, 2019a for discussions of frameworks and dimensions within these). These tools include, the PRagmatic-Explanatory Continuum Indicator Summary (PRECIS) which included 10 domains with a visual scale to represent where the trial fell on each domain (K. E. Thorpe et al., 2009). PRECIS was later revised to PRECIS-2 (Loudon et al., 2015; Loudon et al., 2013b) which further revised the domains but also incorporated a quantification of the degree to which the design reflected the underlying pragmatic orientation of the trial (using a 5-point Likert-style scale). This addition of a score within each domain was consistent with work that had been undertaken in the intervening period and which had sought to apply scoring mechanisms to the original PRECIS domains. The introduction of quantitative assessment was due to noted variation in the application of the original PRECIS tool and was thus sought to standardise the assessment of trial design features by introducing a common scale.

As these tools have evolved, so has their use. In contrast to the initial development of tools for prospective use by trial investigators to evaluate their own trial, tools such as the PRECIS-RT tool (Koppelaar et al., 2011) have been developed and applied in retrospective analyses of trial reports. This has represented a significant change in orientation from prospective assistance for trial investigators to retrospective evaluation by researchers not involved in the original trial. Multiple studies have now sought to apply existing tools retrospectively in order to classify trials as either pragmatic or explanatory (Aves et al., 2017; Devos et al., 2019; Palese et al., 2014; Steel et al., 2017; Yoong et al., 2014). For example, Yoong et al (2014) explored the intervention effect size according to the categorization of a

trial, while Aves et al sought to explore whether designation of the trial may be informative with respect to explaining inconsistent trial results (Aves et al., 2017). As such, despite repeated statements that RCTs sit along an explanatory-pragmatic spectrum (Gartlehner et al., 2006; Neta & Johnson, 2018; Oxman et al., 2009; K. E. Thorpe et al., 2009), or indeed a multi-axial continuum, there are now attempts to retrospectively evaluate and score trials and apply thresholds by which to dichotomise trials as pragmatic or not.

Both of these moves (retrospective assessment of the degree of pragmatism in design features, and overall classification of trial design as pragmatic or explanatory) reflect a change from the original orientation which was to assist trial investigators understand the degree to which their trial design was consistent with the intention of the trial.

### **Why does this matter?**

The importance of the co-consideration of intent and design is underscored by the failure of many trials to predict the actual effectiveness of an intervention in later clinical practice (Zwarenstein & Treweek, 2009), in part due to the disconnect between their intent and their design. The integration of intent is important because a focus solely of metrics (and subsequent dichotomous categories of pragmatic or explanatory trials based on this) abstracts design from intent and whether a design is ‘fit for purpose’. Indeed, the increasing focus on metrics illustrates the partial adoption of tools such as PRECIS-2; attention has been focused on the production of the scores, or in the case of PRECIS-2, the ‘wheel’, to the detriment of other aspects such as consideration of the PRECIS-2 table which requires investigators to provide a rationale for the design choices within each PRECIS-2 domain (<https://www.precis-2.org/Help/Documentation/ToolkitDownload>). While the PRECIS-2 domain scores can provide a useful, shorthand assessment of a trial’s pragmatism along various dimensions, the explicit rationales sought by

the PRECIS-2 table offer a far more informative picture of how, and in what respects, the trial investigators considered their trial to be pragmatic.

While one may argue that all trials in which the intent is to generate evidence applicable to a clinical or policy decision should be required to score highly on tools such as PRECIS-2, it pays no consideration to legal requirements, questions of feasibility, nor whether particular elements of design may be more appropriately designed to be more explanatory in order to best provide evidence relevant to the clinical or health policy question at hand.

In order to more completely understand and evaluate the design features of a trial, both as originally designed and subsequently operationalised, the design information must be placed into context. We therefore assert that the PRECIS-2 table, or similar demonstration of design rationale, is integral to evaluating whether the study design is consistent with the intent of the trial.

### **Integrating intent**

It should be remembered that pragmatic RCTs are intended to have their results be *applicable* to clinical or health policy decisions (Maclure, 2009; Zwarenstein & Treweek, 2009). While tools such as PRECIS were developed to assist investigators to evaluate their trial design with respect to the decision that the trial intended to inform, it does not prescribe a specific study design or set of design features. A focus on the retrospective quantitative evaluation and classification based on this score completely ignores whether the trial is consistent with the stated intent, which may reflect more pragmatic or explanatory attitudes. Moreover, the applicability of trial results will depend on the specific context of the trial and transferability of the trial results rather than an abstract score of the design features (Pawson, 2019b). Rather, we reiterate that the research design should be predicated on the research question which derives from the decision that the trial seeks to inform.

To further illustrate our position, Table 1 provides examples of studies that clearly articulate the intention of the trial.

[TABLE 1 ABOUT HERE]

### **Conceptual and practical difficulties in retrospectively applying thresholds based on PRECIS-2**

In addition to the above noted need to consider the intent of the trial, we see several conceptual and practical difficulties in the retrospective evaluation and classification of RCTs, namely: inconsistency and lack of consensus regarding thresholds by which to establish what constitutes a pragmatic trial, epistemic uncertainty regarding the relevant contribution of specific domains to the overall evaluation of the trial, and; practical limitations due to noted variation in assessment and incompleteness of reporting.

[TABLE 2 ABOUT HERE]

First, proposed thresholds used to categorise trials have varied, with no consensus about the thresholds to use (Aves et al., 2017; Devos et al., 2019; Sajobi et al., 2018; Steel et al., 2017; Yoong et al., 2014). Yoong and colleagues (Yoong et al., 2014), created an ordered set of categories (explanatory, a combination of pragmatic and explanatory, or pragmatic) with thresholds based on the average of scores across the nine PRECIS-2 domains. Sajobi et al (2018) used the same scoring mechanism as Yoong and colleagues (average score across PRECIS-2 domains). Rather than an ordered set of categories they applied a threshold to dichotomise the classification (see Table 2). In each case, the proposed thresholds lack clear rationales or justifications for their choice, leading to incompatibility between studies. Further, there is epistemic uncertainty as to the relevant weight that should be ascribed to individual design features, that is, each domain of PRECIS-2 (Dekkers et al., 2017; Pawson, 2019a). This question of weighting was raised by Koppenaal et al., (2011) in their work to develop PRECIS-RT, where they noted:

“Eligibility criteria are likely to always be crucial, but the flexibility of the comparison intervention, for example, may sometimes be less important. We do not have a clear answer to this problem, especially because the best weighting of the domains could depend on the situation.” (Koppenaar et al., 2011)

This again emphasises that the importance of design features should be evaluated with respect to the stated intent of the trial; the weighting may be dependent on the situation, that is the intent of the trial or the decisions to be informed. For example, criteria such as participant eligibility will likely be relevant to all studies, yet the flexibility of the comparison intervention, or organisation and structure (if part of the intervention) may vary in importance (Dal-Ré et al., 2018; Koppenaar et al., 2011). Yet in most studies that seek to apply quantitative scores to determine whether a trial is pragmatic or not, the relative contribution of each domain to the overall score is not discussed (Luoma et al., 2017; Sanchez et al., 2013; Sepehrvand et al., 2019; Witt et al., 2012). We believe that the relative contributions that individual domains *should* contribute toward the overall assessment of the trial – including the extent to which the domain-related design features appropriately reflect the intent of the trial – requires further conceptual, and potentially empirical, consideration.

Second, and on a more practical note, the ability to evaluate and classify trials retrospectively is highly varied. As Pawson notes, a key issue is whether different observers, from different backgrounds and interested in different conditions, will view the dimensions in the same manner and be equally calibrated so as to come to the same conclusion regarding the pragmatism of a trial (Pawson, 2019a). In the development of PRECIS-2, Loudon et al. (Loudon et al., 2017) conducted an inter-rater reliability study regarding the retrospective application of the PRECIS-2 tool. They noted that seven of nine domains had an intraclass correlation coefficient (ICC) over 0.65 but that two (flexibility-adherence, and recruitment) had lower ICCs and wide confidence intervals (Loudon et al., 2017). This variation in domain scores has been found across empirical studies of retrospective evaluation (Gaglio et al., 2014; Glasgow et al., 2012;

Loudon et al., 2017). Moreover, Loudon et al., (Loudon et al., 2017) found that discriminant validity of the PRECIS-2 scores was modest, further supporting the argument to say that such a tool is of limited use with respect to retrospectively evaluating trials as pragmatic or explanatory. Again Pawson, as we concur, states that a challenge here is that such measurement itself is indirect (Pawson, 2019a). A score on a tool such as PRECIS-2 is not an intrinsic metric — rather, each domain score can be thought of as an indicator of an underlying latent construct requiring judgement or interpretation which may differ between individuals.

Finally, the scoring of trials retrospectively is further hampered by missing data in reports, a topic that has been repeatedly highlighted by authors attempting to retrospectively apply scoring mechanisms (Aves et al., 2017; Koppelaar et al., 2011; Loudon et al., 2017; Yoong et al., 2014). The variation in the retrospective scoring of trials, together with noted prevalence of missing data, illustrates that quantifying the degree of pragmatism within a trial based purely on reported design features is both variable and subject to how well the trial is reported.

It seems premature, therefore, to advocate for the use of retrospective evaluation and categorisation of trials as pragmatic or not given the noted variation in assessment, conceptual uncertainty regarding weighting of domains, and the absence of consensus regarding where thresholds should be placed (if at all).

### **Practical implications of our proposal**

We propose that trial investigators should explicitly state their intent through explicit discussion of the decision(s) to which the trial is intended to provide applicable data. This should be done prospectively as they develop the design of the trial. The study design should then flow from this intent and thus the design of a trial should be judged in conjunction with the stated intent. Further, we suggest that in addition to a formal statement regarding the decision that the trial results should be applicable to,

investigators should indicate the relative importance of design features to the decision at hand. For example, in the context of pragmatic RCTs comparing vaccine dosing schedules or storage (Juan-Giner et al., 2014; Neuzil et al., 2011), the primary outcome may be immunogenicity based on antibody levels. This may, by examination of the PRECIS-2 guidance, be scored as a very explanatory outcome. However, given the prominence of this outcome within policy decision-making, this outcome choice may be appropriate and as such a lower PRECIS-2 score may carry less weight with respect to the overall assessment of the trial. Similarly, the importance of pragmatism in the analytic approach (for example the use of intention to treat analysis and per protocol analysis) may vary depending on the question being addressed (Murray et al., 2018; Murray et al., 2019). In Supplementary Material S1 we propose an enhanced version of the existing PRECIS-2 toolkit table which incorporates these suggestions.

Despite our suggestion, we acknowledge that there is an implementation gap between the completion of a trial and the impact, if any, of that trial on clinical practice or policy. This does not, we believe, change the need for the reports of pragmatic trials to be accessible and understandable. We believe our proposal would provide the contextual information to evaluate the appropriateness of the study design in relation to the decision that the trial results were intended to be applicable. This is not a new requirement and is explicitly included within the CONSORT extension for pragmatic trials which states:

“Users of pragmatic trial reports seek to solve a health or health service problem in a particular setting. The problem at which the intervention is targeted should thus be described. This enables readers to understand whether the problem confronting them is similar to the one described in the trial report, and thus whether the study is relevant to them. Ideally, *the report should state that the trial is pragmatic in attitude (and why) and explain the purpose of the trial in relationship to the decisions that it is intended to inform and in which settings.*” (Zwarenstein et al., 2008) (emphasis added)

We suggest that this should not be an aspirational ideal but a necessary component, and that both the trial protocol and original trial report should explicitly detail how pragmatic design features reflect the needs of the trial in relation to the decision that the trial is intended to inform. We suggest that a completed version of our enhanced PRECIS-2 table could be included as supplementary material for trials with a pragmatic intent. Further, recent analysis of self-identified pragmatic trials supported by the US National Institutes of Health illustrates how initial trial designs may be required to adapt when a trial begins implementation.(Johnson et al., 2016), suggesting that clear reporting is needed at both the concept or protocol stage and upon trial completion. Understanding how the implemented trial is consistent or divergent from the initial intent will again enhance understand of the degree to which the findings of the implemented trial are relevant to the decision that the trial results were intended to inform. In short, it allows for the explicit examination of whether the final trial design was in-keeping with the intent and thus ‘fit for purpose.’

We thus concur with Dal-Ré and colleagues (Dal-Ré et al., 2018) that transparent and complete reporting of trials is essential to their proper evaluation, and that adherence to the proposals within the CONSORT extension for pragmatic trials should be promoted (Zwarenstein et al., 2008). This will not just benefit researchers and knowledge users with respect to making assessments regarding the intent of the trial and the degree to which this reflects a pragmatic orientation, but it will also facilitate an understanding of appropriateness of the consequent trial design.

## **Conclusion**

In conclusion, we argue against the quantitative evaluation and categorisation of trials as pragmatic or explanatory *solely* on the basis of design elements, such as those provided in PRECIS-2. Moreover, we suggest that it is the responsibility of investigators to state the relative emphasis that should be placed on specific design domains. However, there remains a need for conceptual and empirical study of the relevant contributions that different domains make toward the overall degree of pragmatism within a trial,

and under what circumstances the weights of the domains may differ. We contend that existing examples of dichotomisation are conceptually problematic but also raise practical concerns, and a more appropriate response requires one to evaluate the overall trial design (as articulated through a complete PRECIS-2 assessment or similar) in the context of the decision the trial was intended to inform. We believe that a focus on the integration of intent and design is more in keeping with the original work with the original work by Schwartz and Lellouch. When coupled with more complete and transparent reporting this will create less arbitrary classification of trials.

### **ACKNOWLEDGEMENTS**

This work is supported by the Canadian Institutes of Health Research through the Project Grant competition (competitive, peer-reviewed), award number PJT-153045.

### **AUTHOR CONTRIBUTIONS**

**Stuart G Nicholls, Merrick Zwarenstein, Spencer P Hey, Bruno Giraudeau, Marion K Campbell, Monica Taljaard:** All authors contributed to the writing, reviewing and editing of the manuscript.

## References

- Alphs, L.D., & Bossie, C.A. (2016). ASPECT-R—A Tool to rate the pragmatic and explanatory characteristics of a clinical trial design. *Innovation in Clinical Neuroscience*, 13, 15-26.
- Aves, T., Allan, K.S., Lawson, D., Nieuwlaat, R., Beyene, J., & Mbuagbaw, L. (2017). The role of pragmatism in explaining heterogeneity in meta-analyses of randomised trials: a protocol for a cross-sectional methodological review. *BMJ Open*, 7, e017887.
- Bossie, C.A., Alphs, L., Williamson, D., Mao, L., Kurut, C., & The ASPECT-R Rater Team (2016). Inter-rater reliability assessment of ASPECT-R. *Innovation in Clinical Neuroscience*, 13, 27-31.
- Chalkidou, K., Tunis, S., Whicher, D., Fowler, R., & Zwarenstein, M. (2012). The role for pragmatic randomized controlled trials (pRCTs) in comparative effectiveness research. *Clinical Trials (London, England)*, 9, 436-446.
- Dal-Ré, R. (2019). The misleading use of the term pragmatic in pre-licensing medicine trials. *European Journal of Clinical Pharmacology*, Online First.
- Dal-Ré, R., Janiaud, P., & Ioannidis, J.P.A. (2018). Real-world evidence: How pragmatic are randomized controlled trials labeled as pragmatic? *BMC Medicine*, 16.
- Dekkers, O.M., Bossuyt, P.M., & Vandenbroucke, J.P. (2017). How trial results are intended to be used: is PRECIS-2 a step forward? *J Clin Epidemiol*, 84, 25-26.
- Devos, F., Foissac, F., Bouazza, N., Ancel, P.Y., Treluyer, J.M., & Chappuy, H. (2019). Study characteristics impacted the pragmatism of randomized controlled trial published in nursing: a meta-epidemiological study. *J Clin Epidemiol*, 116, 18-25.
- Ford, I., & Norrie, J. (2016). Pragmatic trials. *N Engl J Med*, 375, 454-463.
- Gaglio, B., Phillips, S.M., Heurtin-Roberts, S., Sanchez, M.A., & Glasgow, R.E. (2014). How pragmatic is it? Lessons learned using PRECIS and RE-AIM for determining pragmatic characteristics of research. *Implementation Science*, 9.

- Gartlehner, G., Hansen, R.A., Nissman, D., Lohr, K.N., & Carey, T.S. (2006). A simple and valid tool distinguished efficacy from effectiveness studies. *J Clin Epidemiol*, 59, 1040-1048.
- Glasgow, R.E., Gaglio, B., Bennett, G., Jerome, G.J., Yeh, H.C., Sarwer, D.B., et al. (2012). Applying the PRECIS criteria to describe three effectiveness trials of weight loss in obese patients with comorbid conditions. *Health Serv Res*, 47, 1051-1067.
- Johnson, K.E., Neta, G., Dember, L.M., Coronado, G.D., Suls, J., Chambers, D.A., et al. (2016). Use of PRECIS ratings in the National Institutes of Health (NIH) Health Care Systems Research Collaboratory. *Trials*, 17, 32.
- Juan-Giner, A., Domicent, C., Langendorf, C., Roper, M.H., Baoundoh, P., Fermon, F., et al. (2014). A cluster randomized non-inferiority field trial on the immunogenicity and safety of tetanus toxoid vaccine kept in controlled temperature chain compared to cold chain. *Vaccine*, 32, 6220-6226.
- Koppelaar, T., Linmans, J., Knottnerus, J.A., & Spigt, M. (2011). Pragmatic vs. explanatory: an adaptation of the PRECIS tool helps to judge the applicability of systematic reviews for daily practice. *J Clin Epidemiol*, 64, 1095-1101.
- Loudon, K., Treweek, S., Sullivan, F., Donnan, P., Thorpe, K.E., & Zwarenstein, M. (2015). The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ*, 350, h2147.
- Loudon, K., Zwarenstein, M., Sullivan, F., Donnan, P., & Treweek, S. (2013a). Making clinical trials more relevant: improving and validating the PRECIS tool for matching trial design decisions to trial purpose. *Trials*, 14, 115.
- Loudon, K., Zwarenstein, M., Sullivan, F., Donnan, P., & Treweek, S. (2013b). PRECIS-2: a tool to improve the applicability of randomised controlled trials. *Trials*, 14, O28.
- Loudon, K., Zwarenstein, M., Sullivan, F.M., Donnan, P.T., Gagyor, I., Hobbelen, H., et al. (2017). The PRECIS-2 tool has good interrater reliability and modest discriminant validity. *J Clin Epidemiol*, 88, 113-121.

- Luoma, K.A., Leavitt, I.M., Marrs, J.C., Nederveld, A.L., Regensteiner, J.G., Dunn, A.L., et al. (2017). How can clinical practices pragmatically increase physical activity for patients with type 2 diabetes? A systematic review. *Transl Behav Med*, 7, 751-772.
- Maclure, M. (2009). Explaining pragmatic trials to pragmatic policy-makers. *CMAJ*, 180, 1001-1003.
- Murray, E.J., Caniglia, E.C., Swanson, S.A., Hernández-Díaz, S., & Hernán, M.A. (2018). Patients and investigators prefer measures of absolute risk in subgroups for pragmatic randomized trials. *Journal of Clinical Epidemiology*, 103, 10-21.
- Murray, E.J., Swanson, S.A., Young, J., & Hernán, M.A. (2019). Guidelines for estimating causal effects in pragmatic randomized trials. In B.U.S.o.P. Health (Ed.). Boston, USA.
- Neta, G., & Johnson, K.E. (2018). Informing real-world practice with real-world evidence: the value of PRECIS-2. *BMC Medicine*, 16.
- Neuzil, K.M., Canh, D.G., Thiem, V.D., Janmohamed, A., Huong, V.M., Tang, Y., et al. (2011). Immunogenicity and reactogenicity of alternative schedules of HPV vaccine in Vietnam. A cluster randomized noninferiority trial. *JAMA*, 305, 1424-1431.
- Oxman, A.D., Lombard, C., Treweek, S., Gagnier, J.J., Maclure, M., & Zwarenstein, M. (2009). A pragmatic resolution. *J Clin Epidemiol*, 62, 495-498.
- Palese, A., Bevilacqua, M.G., Dante, A., & Group, E.P. (2014). Do randomized controlled nursing trials have a pragmatic or explanatory attitude? Findings from the Pragmatic-Explanatory Continuum Indicator Summary (PRECIS) tool exercise. *J Nurs Res*, 22, 216-220.
- Patsopoulos, N.A. (2011). A pragmatic view on pragmatic trials. *Dialogues in Clinical Neuroscience*, 13, 217-224.
- Pawson, R. (2019a). The "pragmatic trial": An essentially contested concept? *J Eval Clin Pract*, 1-12.
- Pawson, R. (2019b). Pragmatic trials and implementation science: grounds for divorce? *BMC Med Res Methodol*, 19, 176.
- Roland, M., & Torgerson, D.J. (1998). Understanding controlled trials: What are pragmatic trials? *BMJ*, 316, 285-285.

- Sajobi, T.T., Li, G., Awosoga, O., Wang, M., Menon, B.K., Hill, M.D., et al. (2018). A comparison of meta-analytic methods for synthesizing evidence from explanatory and pragmatic trials. *Syst Rev*, 7, 19.
- Sanchez, M.A., Rabin, B.A., Gaglio, B., Henton, M., Elzarrad, M.K., Purcell, P., et al. (2013). A systematic review of eHealth cancer prevention and control interventions: new technology, same methods and designs? *Transl Behav Med*, 3, 392-401.
- Schwartz, D., & Lellouch, J. (1967). Explanatory and pragmatic attitudes in therapeutical trials. *Journal of Chronic Diseases*, 20, 637-648.
- Sedgwick, P. (2014). Explanatory trials versus pragmatic trials. *BMJ*, 349, g6694.
- Sepehrvand, N., Alemayehu, W., Das, D., Gupta, A.K., Gouda, P., Ghimire, A., et al. (2019). Trends in the Explanatory or Pragmatic Nature of Cardiovascular Clinical Trials Over 2 Decades. *JAMA Cardiol*.
- Steel, A., Sundberg, T., Reid, R., Ward, L., Bishop, F.L., Leach, M., et al. (2017). Osteopathic manipulative treatment: A systematic review and critical appraisal of comparative effectiveness and health economics research. *Musculoskelet Sci Pract*, 27, 165-175.
- Thorpe, K.E., Oxman, A.D., Treweek, S., & Furberg, C.D. (2010). Pragmatic trials are randomized and may use a placebo. *Journal of Clinical Epidemiology*, 63, 694-695.
- Thorpe, K.E., Zwarenstein, M., Oxman, A.D., Treweek, S., Furberg, C.D., Altman, D.G., et al. (2009). A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *Journal of Clinical Epidemiology*, 62, 464-475.
- Tosh, G., Soares-Weiser, K., & Adamns, C.E. (2011). Pragmatic vs explanatory trials: the Pragmascope tool to help measure differences in protocols of mental health randomized controlled trials. *Dialogues in Clinical Neuroscience*, 13, 209-215.
- Wieland, L.S., Berman, B.M., Altman, D.G., Barth, J., Bouter, L.M., D'Adamo, C.R., et al. (2017). Rating of Included Trials on the Efficacy-Effectiveness Spectrum: development of a new tool for systematic reviews. *Journal of Clinical Epidemiology*, 84, 95-104.

- Witt, C.M., Manheimer, E., Hammerschlag, R., Ludtke, R., Lao, L., Tunis, S.R., et al. (2012). How well do randomized trials inform decision making: systematic review using comparative effectiveness research measures on acupuncture for back pain. *PLoS One*, 7, e32399.
- Yoong, S.L., Wolfenden, L., Clinton-McHarg, T., Waters, E., Pettman, T.L., Steele, E., et al. (2014). Exploring the pragmatic and explanatory study design on outcomes of systematic reviews of public health interventions: a case study on obesity prevention trials. *J Public Health (Oxf)*, 36, 170-176.
- Zwarenstein, M., & Treweek, S. (2009). What kind of randomized trials do we need? *CMAJ*, 180, 998-1000.
- Zwarenstein, M., Treweek, S., Gagnier, J.J., Altman, D.G., Tunis, S., Haynes, B., et al. (2008). Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ*, 337, a2390.



**Table 1: Published trials with clear statement of trial intent**

Trial name and reference	Text demonstrating intent
Cox, et al., (2014). "Impact of Xpert MTB/RIF for TB diagnosis in a primary care clinic with high TB and HIV prevalence in South Africa: a pragmatic randomised trial." PLoS Med 11(11): e1001760.	"While significant data exist on the specificity and sensitivity of Xpert testing [6], there are limited data available on the impact of implementing Xpert on health outcomes in routine programmatic settings [7,8]. As a result, there remains controversy as to how Xpert should be implemented in different health systems and who should be tested. <i>We aimed to assess the impact of using Xpert for TB diagnosis on yield of TB cases and the timing of TB treatment initiation in a large primary health care clinic, through a pragmatic randomised controlled trial</i> "
Witt, et al., (2015). "Effectiveness of an additional individualized multi-component complementary medicine treatment on health-related quality of life in breast cancer patients: a pragmatic randomized trial." Breast Cancer Res Treat 149(2): 449-460.	" [...] In January 2010, the regional public health system in South Tyrol (Italy) established a service for CM at the Merano Hospital planned to run for 2 years, and aiming to improve quality of life in cancer patients and patients with chronic conditions. <i>The present study was initiated by the regional public health system to inform the decision whether or not the service should be maintained after December 2012. Therefore, the primary aim of this study was to evaluate the effectiveness of an additional, individualized, multi-component CM treatment offered at the Merano Hospital compared to usual care only on health-related quality of life in patients with breast cancer.</i> "
Anderson, et al., (2018). "Randomised controlled trial to assess the impact of a lifestyle intervention (ActWELL) in women invited to NHS breast screening." <i>BMJ Open</i> 8(11): e024136.	"The current study is designed to assess the effectiveness of a community-based, personalised, minimal contact weight management programme in women with a BMI >25 kg/m <sup>2</sup> attending routine breast cancer screening clinics. The intervention programme is a collaboration between the charity Breast Cancer Now (BCN), NHSSBSP, local authority leisure centres and academic partners. This work is the first time that a cancer charity has offered volunteer capacity for cancer prevention action on weight management and offers significant potential to address gaps in public health efforts. <i>The design is pragmatic to increase the relevance of the findings to policymakers, women eligible for breast screening and health professionals (see online supplementary appendix 1).</i> " p2-3
Than, et al., (2016). "Effectiveness of EDACS Versus ADAPT Accelerated Diagnostic Pathways for Chest Pain: A Pragmatic Randomized Controlled Trial Embedded Within Practice." <i>Ann Emerg Med</i> 68(1): 93-102 e101.	"Clinicians do not always adhere to clinical pathways or guidelines as expected, and it is important to determine whether the EDACS-ADP would work within a clinical pathway implemented into daily hospital care when the attending clinician has final decision-making authority. We therefore designed a trial to test for the existence and size of any beneficial effect of using the EDACS-ADP in routine clinical care. We tested the null hypothesis that there was no difference in using the EDACS-ADP to classify patients to low-risk category and early discharge from the ED than using the modified ADAPT-ADP."
Butler et al., (2020). "Oseltamivir plus usual care versus usual care for influenza-like illness in primary care: an open-label, pragmatic, randomised controlled trial." <i>The Lancet</i> 395(10217): 42-52.	"We deliberately chose to do an open-label trial in the context of everyday practice, because effect sizes identified by placebo-controlled, efficacy studies with tight inclusion criteria might not be reproduced in routine care. We also wished to estimate time to patient reported recovery from the addition of an antiviral agent to usual care rather than benefit from oseltamivir treatment compared with placebo [...] This pragmatic, open trial design makes our findings likely to reflect real world effects in primary care, because knowledge of what medication one is taking could affect subsequent help seeking and health behaviour and use of symptomatic medications. uncomplicated influenza found a similar effect to our study overall and observed reductions in the duration of symptoms and virus shedding even when treatment was started more than 48 h after illness onset."

**TABLE 2: Studies applying thresholds to the PRECIS-2 domain scores**

Author	Rationale for categorisations	Scoring mechanism	Rationale for threshold level(s)
Yoong et al., 2014	“Given that findings from pragmatic trials are more likely to closely approximate the public health impacts caused by an intervention if investments were made to introduce it into the community, examining intervention outcomes based on trial design is likely to provide more useful information for assessing the transferability of public health initiatives	Studies were classified as Pragmatic (average score on PRECIS-2 >2.2); Combination pragmatic/explanatory (average score >1.7 - ≤ 2.2 ), or Explanatory (average score ≤ 1.7 ).	Not given
Aves et al 2017	“If heterogeneity is substantial, due to the degree of pragmatism, it might not be appropriate to pool data from pragmatic and explanatory trials. The use of the PRECIS-2 tool could provide important information for authors of systematic reviews with regards to pooling data from primary RCTs based on the degree of pragmatism.”	Proposed to use the same cut offs as Yoong et al., (2014)	Not given
Steel et al 2017	Not given	“Articles reporting comparative effectiveness research were scored using the PRagmatic-Explanatory Continuum Indicator Summary-2 (PRECIS-2) tool (Loudon et al., 2015) and independently categorised as employing either an explanatory or observational design by two investigators”	Not given
Sajobi et al. 2018	“Although modern meta-analytic methods such as mixture metaression and robust meta-analytic methods have been developed to pool evidence from heterogeneous populations [26–28], there is limited application of these methods and incorporation of PRECIS ratings in synthesizing evidence from explanatory and pragmatic trials.”	Explanatory if PRECIS-2 average score <3.0, Pragmatic if ≥ 3.0.	Not given
Dal-Ré et al., 2018	“Trials with features that defy pragmatism have been labeled as pragmatic in all types of journal, including major general medical journals such as BMJ and Annals of Internal Medicine. These cases exemplify how the use of the term “pragmatic” needs better standardization.”	Extremely pragmatic: of all 9 domains score ≥4 Pragmatic: scores ≥4 in 4–5 domains, provided the scores of the remaining domains are 3 Exception: highly pragmatic trials where the intervention is how care is organized (and hence the ‘organisation’ domain score will be in the explanatory extreme [12]), could be labeled as pragmatic since this is just an explicit feature of the intervention.	Not given

**Supplementary Material S1. Chronological listing (by year of publication) of selected assessment frameworks and tools developed to capture the domains upon which trials may be more or less pragmatic**

Author	Year	Tool	Purpose/Rationale	Details	Scoring
Gartlehner et al.(Gartlehner et al., 2006)	2006	Not formally named, but described as a tool to distinguish effectiveness from efficacy trials	“Distinguishing between efficacy and effectiveness contributes an important aspect to analyzing any body of clinical evidence [...] we propose and test seven hallmarks of study design to create a tool that can <i>help researchers and those producing systematic reviews</i> , as well as clinicians who are interested in the generalizability of study results, to distinguish more readily and more consistently between efficacy and effectiveness studies.”	7 items: Populations in primary care, Less stringent eligibility criteria, Health outcomes, Long study duration with clinically relevant treatment modalities, Assessment of adverse events, Adequate sample size to assess a minimally important difference from a patient perspective, Intention To Treat (ITT) analysis	Items rated as binary Yes/No
Thorpe, et al(K. E. Thorpe et al., 2009)	2009	PRECIS (pragmatic-explanatory continuum indicator summary.)	“The primary aim of this tool is to help trialists to assess the degree to which design decisions align with the trial’s stated purpose (decision-making vs. explanation). Our tool differs therefore from that of Gartlehner et al. in that it is intended to <i>inform trial design</i> rather than provide a method of classifying trials for the purpose of systematic reviews.”	10 domains: Participant eligibility criteria, Experimental intervention flexibility, Experimental intervention practitioner expertise, Comparison intervention, Comparison intervention practitioner expertise, Follow up intensity, Primary trial outcome, Participant adherence to the “prescribed” intervention, Practitioner adherence to study protocol, Analysis of primary outcome.	Visual ‘hub and spoke’ diagram for the domains, but no formal scoring.
Tosh G, et al.(Tosh et al., 2011)	2011	PRAGMASCOPE	“The main goal of this study is to adapt the instrument described by Thorpe et al (PRECIS) to assist researchers in <i>making those judgments in the protocol stage of RCTs</i> in mental health (the Pragmascope tool).”	Includes the 10 PRECIS Domains: Participant eligibility criteria, Experimental intervention flexibility, Experimental intervention practitioner expertise, Comparison intervention, Comparison intervention practitioner expertise, Follow up intensity, Primary trial outcome, Participant adherence to the “prescribed” intervention, Practitioner adherence to study	Each domain is scored from 1 (most explanatory) to 5 (most pragmatic). Scores summed and broad grouping applied: “scores was of 0 to 30 for an explanatory study investigating whether the experimental intervention will work in ideal circumstances and a total score >35 for a more pragmatic study focusing mostly on whether, in routine practice, an

				protocol, Analysis of primary outcome.	intervention has a meaningful effect. A total score between 31 and 39 were interpreted as an interim where trial design balances pragmatic and explanatory domains.” Later noted as groupings: 0-30, 31-39, 40-50
Koppenaar et al.(Koppenaar et al., 2011)	2011	PRECIS-RT (PRECIS-Review Tool)	“[...] because the PRECIS tool results in a figure for each trial, it is not possible to assess a review. Therefore, in this article we propose a modification of the PRECIS tool <i>so that it can be used to judge a systematic review</i> . This serves two purposes. First, calculating a numeric score makes it possible to score a systematic review and place it on the pragmatic-explanatory continuum. Second, the separate trials in the review can be scored. This may help in selecting the trial that is the most pragmatic, assuming that the results of that trial are the most relevant for policy makers.”	Applied the PRECIS domains within two systematic reviews. For each review. Each RCT within the systematic review is scored on the 10 domains; an average score per RCT is calculated; a domain average across the review calculated, and; a total average for the systematic review can be calculated.	Each domain is scored 1 (extreme explanatory study) to 5 (extreme pragmatic study). Scores also transformed to percentages where 0% represents an extremely explanatory study and 100% represents an extremely pragmatic study.
Loudon et al..(Loudon et al., 2015)	2015	PRECIS-2	“While acknowledging the usefulness of PRECIS, these latter authors have identified weaknesses, including unclear face validity and inter-rater reliability, the lack of a scoring system, redundancy in some PRECIS domains, and the need for more guidance on how to use the tool.”	Nine domains: Eligibility, Recruitment, Setting, Organisation, Flexibility (delivery), Flexibility (adherence), Follow-up, Primary outcome, Primary analysis.	Each domain can be scored from 1 (very explanatory) to 5 (very pragmatic)
Alphs LD, Bossie CA.(Alphs & Bossie, 2016) and Bossie et al..(Bossie et al., 2016)	2016	ASPECT-R	“Janssen Pharmaceuticals, Inc. has adapted ideas from these instruments to build a still more versatile instrument. The number of domains has been reduced to six that are specifically explanatory-pragmatic spectrum. Domains identified as	Developed by Janssen Pharmaceuticals, Inc (2014). Includes 6 domains; Participant eligibility criteria, Intervention flexibility, Medical practice setting/practitioner expertise, Follow up intensity and	Each item score from 0 (extremely explanatory) to 6 (extremely pragmatic).

			<p>redundant and domains focused on measures of study quality have been eliminated.”(Alphs &amp; Bossie, 2016)</p> <p>“As opposed to the 10-domains of the PRECIS and Pragmascope instruments, the ASPECT-R tool assesses six study design domains, with these domains specifically related to the explanatory: pragmatic spectrum. The four domains excluded when developing the ASPECT-R tool were those considered to be redundant or more focused on measures of study quality [...] The ASPECT-R tool is <i>considered useful in the study design stage as well as to assess the explanatory versus pragmatic nature of published trials.</i>”(Bossie et al., 2016)</p>	<p>duration, Outcome(s) and, Participant adherence.</p>	
Wieland et al.(Wieland et al., 2017)	2017	Rating of Included Trials on the Efficacy-Effectiveness Spectrum (RITES)	<p>“PRECIS and PRECIS-2 were developed to inform choices during the trial design phase, rather than to <i>assess the characteristics of trial evidence retrospectively from the publication of the trial.</i> They assume detailed familiarity with available design options at the time that the trial is being designed, and this information may not be available in the report of a completed trial. In addition, PRECIS-2 assesses nine trial domains which may limit the practicality for use on the often substantial number of trials included in a <i>systematic review.</i>”</p>	<p>Four domains: participants’ characteristics, trial setting, flexibility of intervention(s), and clinical relevance of experimental and comparison intervention(s).</p>	<p>Each domain is rated on a five-point scale from a strong emphasis on efficacy to a strong emphasis on effectiveness</p>

## Supplementary Material S2: Proposed Enhanced PRECIS-2 table

### **Step 1:** Why are you doing your trial?

The first step is to be clear why you are doing your trial. Are you:

1. Aiming to take an explanatory approach to answer the question ‘Can this intervention work under ideal conditions?’
2. Aiming to take a pragmatic approach and answer the question ‘Does this intervention work under usual conditions?’

If the latter then specify the decision to be informed, noting the particular stakeholder perspective taken. Is this a clinical decision – for example, which treatment should be selected? – or is it perhaps a health policy decision about what interventions to provide or fund?

**Step 2:** Consider your trial design choices for each of the nine PRECIS-2 domains. Using the table score each domain on a score of 1 to 5 where these correspond to:

1. Very explanatory
2. Rather explanatory
3. Equally pragmatic/explanatory
4. Rather pragmatic
5. Very pragmatic

**Step 3:** For each domain explain how the design chosen, and the degree of pragmatism, relates to the decision to which the trial results should be applied. As part of this investigators should discuss each domain in terms of its relative importance to the decision that is to be informed. For example, a physiological outcome may be chosen (and may be scored as “1.Very explanatory”), but this may be the main clinical outcome used in practice and so may be an appropriate outcome and relevant to the decision being informed. Similarly, the importance of pragmatism in the analytic approach (for example the use of intention to treat analysis and per protocol analysis) may vary depending on the question being addressed.

The table should be used in conjunction with the PRECIS-2 “wheel” or instead of the wheel to give rationale for scores. You can use this to assist discussion with trial collaborators.

Description of decision that the trial is intended to inform			
	Domain	Score	Rationale and description of the design choice in relation to the decision that the trial is intended to inform. Each domain should also be discussed in terms of its relative importance to the decision that is to be informed
1	Eligibility Criteria		
2	Recruitment Path		
3	Setting		
4	Organisation intervention		
5	Flexibility of experimental intervention(s) (delivery)		
6	Flexibility of experimental intervention(s) (adherence)		
7	Follow up		
8	Primary outcome		
9	Primary analysis		

**The 9 PRECIS-2 domains are:**

**Eligibility** –to what extent are the participants in the trial similar to those who would receive this intervention if it was part of usual care? For example, score 5 for very pragmatic criteria essentially identical to those in usual care; score 1 for a very explanatory approach with lots of exclusions (e.g. those who don't comply, respond to treatment, or are not at high risk for primary outcome, are children or elderly), or uses many selection tests not used in usual care.

**Recruitment** - how much extra effort is made to recruit participants over and above what that would be used in the usual care setting to engage with patients? For example, score 5 for very pragmatic recruitment through usual appointments or clinic; score 1 for a very explanatory approach with targeted invitation letters, advertising in newspapers, radio plus incentives and other routes that would not be used in usual care.

**Setting** – how different is the setting of the trial and the usual care setting? For example, score 5 for a very pragmatic choice using identical settings to usual care; score 1, for a very explanatory approach with only a single centre, or only specialised trial or academic centres.

**Organisation** – how different are the resources, provider expertise and the organisation of care delivery in the intervention arm of the trial and those available in usual care? For example, score 5 for a very pragmatic choice that uses identical organisation to usual care; score 1 for a very explanatory approach if the trial increases staff levels, gives additional training, require more than usual experience or certification and increase resources.

**Flexibility (delivery)** – how different is the flexibility in how the intervention is delivered and the flexibility likely in usual care? For example, score 5 for a very pragmatic choice with identical flexibility to usual care; score 1 for a very explanatory approach if there is a strict protocol, monitoring and measures to improve compliance, with specific advice on allowed cointerventions and complications.

**Flexibility (adherence)** - how different is the flexibility in how participants must adhere to the intervention and the flexibility likely in usual care? For example, score 5 for a very pragmatic choice involving no more than usual encouragement to adhere to the intervention; score 1 for a very explanatory approach that involves exclusion based on adherence, and measures to improve adherence if found wanting. In some trials eg surgical trials where patients are being operated on or Intensive Care Unit trials where patients are being given IV drug therapy, this domain is not applicable as there is no compliance issue after consent has been given, so this score should be left blank.

**Follow-up** - how different is the intensity of measurement and follow-up of participants in the trial and the likely follow-up in usual care? For example, score 5 for a very pragmatic approach with no more than usual follow up; score 1 for a very explanatory approach with more frequent, longer visits, unscheduled visits triggered by primary outcome event or intervening event, and more extensive data collection.

**Primary outcome** – to what extent is the trial's primary outcome relevant to participants? For example, score 5 for a very pragmatic choice where the outcome is of obvious importance to participants; score 1 for a very explanatory approach using a surrogate, physiological outcome, central adjudication or use assessment expertise that is not available in usual care, or the outcome is measured at an earlier time than in usual care.

**Primary analysis** – to what extent are all data included in the analysis of the primary outcome? For example, score 5 for a very pragmatic approach using intention to treat with all available data; score 1 for a very explanatory analysis that excludes ineligible post-randomisation participants, includes only completers or those following the treatment protocol.

<https://www.precis-2.org/Help/Documentation/ToolkitDownload>