Journal of Clinical Epidemiology 126 (2020) 193–201

Journal of Clinical Epidemiology

PRECIS SERIES

The design can limit PRECIS-2 retrospective assessment of the clinical trial explanatory/pragmatic features

Rafael Dal-Ré[a,*], Anthonius de Boer[b], Stefan K. James[c]

[a]*Epidemiology Unit, Health Research Institute-Fundación Jiménez Díaz University Hospital, Universidad Autónoma de Madrid, Avda Reyes Católicos 2, E-28040 Madrid, Spain*

[b]*Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Faculty of Science, Utrecht University, PO Box 80082, 3508TB Utrecht, the Netherlands*

[c]*Department of Medical Sciences and Uppsala Clinical Research Center, Uppsala University, Dag Hammarsk€olds v€ag 14B, SE-752 37 Uppsala, Sweden*

Accepted 18 March 2020; Published online 21 April 2020

Keywords: Clinical trials; PRECIS-2; Pragmatic trials; Explanatory trials; Assessment; Blinding; Placebo; Reporting bias

The pragmatic-explanatory continuum indicator summary (PRECIS-2) tool was originally developed to help investigators to design their trials based on the intended purposes [1]. If investigators primarily wanted to support decision makers, then the randomized controlled trial (RCT) should be conducted resembling the usual situation in which the decision between two options is to be made—this is the pragmatic RCT (pRCT). If investigators wanted to explain the way the two options work, then the RCT is conducted under standardized conditions—this is the explanatory RCT. A typical pRCT is an open-label multicenter trial, assessing the comparative effectiveness of two interventions prescribed to a population of individuals with rather different characteristics and with the same number of visits and procedures as in normal clinical practice. There are few pRCTs conducted, and pRCTs with medicines are rarely performed [2,3]. A typical explanatory RCT is a double-blind phase 3 trial comparing the efficacy of an investigational medicine with that of a placebo or an active medication, in which strict eligibility criteria lead to a recruited population of participants of very similar characteristics and where the number of visits and procedures is much higher than in normal clinical practice. Close to 6,000 phase 3 RCTs were registered on ClinicalTrials.gov as ‘‘recruiting’’ or ‘‘active’’ as of November 2019 [4].

1. The PRECIS-2 tool in practice

The PRECIS-2 tool considers nine domains: eligibility, recruitment, setting, organization, flexibility: delivery, flexibility: adherence, follow-up, primary outcome, and primary analysis [1]. All these nine domains should be individually scored from one (very explanatory) to five (very pragmatic); three means equally pragmatic and explanatory. Adding up the scores of the nine domains, the PRECIS-2 total trial score could fall in the explanatory side or in the pragmatic side. The issue is when an RCT can reasonably be considered (and labeled) as pragmatic or as explanatory, apart from the scores of each of the nine domains.

The PRECIS-2 tool is used both prospectively and retrospectively to assess the degree of pragmatism of RCTs. To conduct a correct assessment with the PRECIS-2 tool, investigators should have broad and deep knowledge of all the nine domains. This is easily achieved when investigators prospectively use the tool at the trial design stage or retrospectively when conducted by the same investigators who run the trial. However, when the PRECIS-2 tool is retrospectively used by investigators who *were not* involved in the conduct of the trial, in our opinion, an acceptable assessment of the degree of pragmatism by means of the PRECIS-2 tool will require having access to the research

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of interest: R.D–R. has received honorarium from Palex Medical for giving a lecture on low-risk pragmatic trials, outside from the submitted work; S.K.J. reports grants and personal fees from AstraZeneca, grants from Bayer, grants and personal fees from Boston, grants from Abbott, grants from Jansen, and grants from Med Co, outside the submitted work; A.d.B. has no conflicts of interest.

* Corresponding author. Instituto de Investigación Sanitaria-Hospital Universitario Fundación Jiménez Díaz, Avda Reyes Católicos 2, 28040 Madrid, Spain. Tel.: +34 649410221; fax: +34 914974083.

E-mail address: rafael.dalre@quironsalud.es (R. Dal-Ré).

https://doi.org/10.1016/j.jclinepi.2020.03.027

0895-4356/© 2020 Elsevier Inc. All rights reserved.

What is new?

Key findings

- The pragmatic-explanatory continuum indicator summary (PRECIS-2) tool was designed to prospectively assess the pragmatic/explanatory features of randomized controlled trials (RCTs), by means of evaluating nine different domains. There is an increasing number of studies retrospectively assessing many RCTs of a given specialty, disease, or condition. In these studies, investigators followed the PRECIS-2 tool assessment recommendations, but failed to consider three key design features we believe are essential to bear in mind in retrospective assessment of many trials: masking, the use of a placebo, and the conduct of the trial in a single center. These three features make RCT explanatory irrespective of the PRECIS-2 total score.

What this adds to what was known?

- We assessed three recently published systematic reviews assessing RCTs on cardiovascular diseases and nursery and rheumatoid arthritis, each using PRECIS-2 to measure the pragmatism. In none of them, the three key design features mentioned above were considered in their assessments, so from our perspective, the PRECIS-2 scores obtained were questionable. Furthermore, we flagged two important issues when retrospectively assessing RCTs: what investigators should do when there is no information on one or more domains and when significant reporting bias is present.

What is the implication and what should change now?

- Investigators should bear in mind that the PRECIS-2 tool is useful when there is no masking, no placebo use, and in multicenter RCTs. In the retrospective use of this tool, the scientific community should agree on what approach should be followed regarding the domains with no or insufficient information and when significant reporting bias is present, so the results of future studies on articles of different specialties/indications could be easily compared, something that now is not possible.

ethics committee approved trial protocol (and subsequent amendments)—although, sometimes, even trial protocols poorly describe one or more domains [5]. In addition, investigators should look for other sources of information such as that included in trial registries (both public and private), other information available on the internet (e.g., presentations), and regulatory information for regulated-intervention trials (e.g., European Medicines Agency's (EMA) clinical study report and Food and Drug Administration's (FDA) drug approval packages) [6–8]. Published protocols—that are becoming more popular in recent years—are also useful, but sometimes lack information to appropriately assess certain domains [9,10]. Relying only on the information reported in the articles that describe trial results, will most likely render a PRECIS-2 questionable assessment because the available information is usually very limited in several domains [6,11–16] (Table 1).

It is well accepted that pragmatic and explanatory are the extremes of a continuum; most trials have both pragmatic and explanatory features [1]. There are exceptions such as a phase 1 first-in-human trial in healthy volunteers, which has only explanatory features. In the other extreme, the GISSI trial [17] on the effectiveness of intravenous thrombolytics in acute myocardial infarction was regarded by their investigators to have all nine domains with a score of five, that is, totally pragmatic [18]. However, are there any trial features that should result in a trial being considered as explanatory regardless of the PRECIS-2 total score?

2. Explanatory or pragmatic: consider key design features first, then, if appropriate, use the PRECIS-2 tool

Before retrospectively using the PRECIS-2 tool to assess how pragmatic/explanatory a trial was by investigators who *were not* involved in it, they should answer two critical questions [2]. First, did the trial resemble usual clinical practice? If the answer is clearly no, then the subsequent use of the PRECIS-2 tool has only a marginal academic interest. In fact, as we will show below, the use of the PRECIS-2 tool could be a source of misunderstandings. The second question is as follows: were the results applicable to other usual care settings, that is, were the results ''generalizable''? This is the most important feature of a pRCT [19]. If the answer is clearly no, then we are facing a similar situation as the one previously described. The relevance of RCTs depends on the generalizability (external validity) of the results to patients seen in usual clinical practice and should be reported in a way readers could consider the applicability of the results to their own context [19,20].

Regarding the first question, we argue that single-, double-blind trials and placebo-controlled trials are so far from usual clinical practice, that they cannot be considered (and labeled) as pragmatic—regardless the PRECIS-2 total trial score. In usual clinical practice, once the treatment alternatives have been discussed, the physician will agree with the patient—taking into consideration her values and preferences—the best treatment to be prescribed. Therefore, patients know the specifics of their treatments. In no case, patients agree to be treated with a placebo. We argue

Table 1. Retrospective assessment of the degree of pragmatism of a published trial by means of the PRECIS-2 tool. PRECIS-2 tool domains with lack of information on articles describing trial results

Domain	PRECIS-2 comments [1]	Information lacking in articles [6,11–16]
Eligibility Who is selected to participate in the trial?	- A highly pragmatic RCT will include all possible candidates (e.g., as per product label or anyone who is receiving the intervention in usual care). In general, a trial protocol with broad inclusion criteria and few exclusion criteria.	- The diagnosis of recruited patients is unclear in some articles. - It is very uncommon to have all inclusion and exclusion criteria reported in the article. - Comorbidity of participants and comedication are usually not reported.
Recruitment How are participants recruited into the trial?	- A highly pragmatic RCT will recruit participants as in usual care (e.g., those attending the primary care center or the emergency room in a hospital). Adding other recruitment strategies (e.g., letters or emails to potential participants; media advertising) will make the trial more explanatory. - In regulated-intervention trials, the informed consent process should comply with the regulations (e.g., detailed and lengthy participants' information sheet in prelicensing trials and shorter in postlicensing trials). - Adding any assessment to those conducted in normal clinical practice will make the trial explanatory.	- Many articles do not report patients' selection before randomization. - Many publications do not describe how participants were recruited. They do not commonly mention the use of unusual recruitment strategies. - Most articles do not to provide details on the informed consent process (e.g., no mention on the detailed and lengthy participants' information sheet to comply with the clinical trial regulations), except for stating that participants provided their written informed consent. - All publications normally mention unusual assessments if they provided reported research data. - Articles commonly mention any substudy (e.g., genetic) if they provided reported research data.
Setting Where is the trial being performed?	- Trialists should consider the match between the setting of the trial and the setting where their results are likely to be applied. If these are identical, the trial will be very pragmatic. - Multicenter trials are easily pragmatic because they make it easier to claim that the trial setting matches that to which the results will be applied. - A single-center trial could be regarded as pragmatic only when it is conducted in a very specialized setting, although the results could only be generalizable to a very restricted number of specialized centers.	- Some articles fail to report the countries where they were run. - Sometimes it is unclear the number of centers involved in the trial. - Articles usually don't provide enough information about the rational of the type of sites involved. - Publications not always report the settings where the trials were run. This is almost the rule in multinational trials, in which sites in each country—or even within a country—could be different settings.
Organisation What expertise and resources are needed to deliver the intervention?	- How easy the implementation of the intervention posttrial is. - A pragmatic trial would be conducted in the usual organization of care for the disease/condition under study, using the same health care staff and resources. - If the trial requires to make changes in how care is delivered compared with usual care, the trial becomes more explanatory. For instance, if the trial requires specific training to staff or additional staff to deliver the intervention. - Trials conducted in secondary care (specialists) when in usual practice patients are seen in primary care (general practitioners) are explanatory.	- Characteristics of healthcare settings are usually not described. - Articles usually fail to report the conduct of trials in mixed settings (primary and secondary care sites). - It is common that publications do not provide information regarding trialists expertise, that usually differs between countries and even within the same country (e.g., USA). - Articles not always report the training that eventually trialists received to deliver the interventions.
Flexibility: (intervention) delivery How should the intervention be delivered?	- Investigators should think how the intervention will be implemented posttrial.	- It is fairly common to describe how safety monitoring was conducted, that will allow knowing if it was similar to usual clinical practice or not.

(*Continued*)

Table 1. Continued

Domain	PRECIS-2 comments [1]	Information lacking in articles [6,11–16]
	- In a highly pragmatic trial, the details of how to implement the intervention will match usual care. - Delivery flexibility applies to all intervention arms. - The trial will be more explanatory if providers undertake additional intervention that do not occur in usual care and if there are specific directions for managing side effects.	- In prelicensing medicine trials, the experimental medicine must be dispensed in a noncommercial package containing the warning ''investigational drug -— for clinical trials use only''. - In single-blind (participant) and double-blind trials, participants will have to receive all medications (and placebos) in noncommercial packages. - In many trials, the medicines are delivered in a different setting than in usual practice (e.g., hospital pharmacies instead of in community pharmacies). This is not always reported.
Flexibility: (intervention) adherence What measures are in place to make sure participants adhere to the intervention?	- A highly pragmatic trial would allow for full flexibility in how end-user recipients engage with the intervention. - A trial that lays out methods to monitor and ensure patient compliance is explanatory. - Delivery adherence applies to all intervention arms; if needed, the arms should be scored separately.	- Some articles stated that participants should return the unused medication to monitor the adherence to treatment. This, however, does not always happen.
Follow-up How closely are participants followed up?	- Trials that have no more follow-up than is normal in usual care and have minimal additional data collection would be likely highly pragmatic. - Trials will be more explanatory if follow-up visits are more frequent than under usual care, or patients are contacted if they fail to keep trial appointments, or if more extensive data are collected than would be typical outside the trial.	- Articles usually mention the number of visits conducted. This will allow knowing if the number of visits resembled usual clinical practice, provided investigators know this information in the setting where the trial was conducted. When reviewing trials conducted in many countries, in many occasions, authors cannot know the number of visits that are normal practice in those countries and for many diseases/conditions. - Articles do not always mention the procedures performed in each visit. Again, it is almost impossible that authors could know the usual procedures performed for any visit in a given disease/condition in all the countries where those trials were conducted. - Publications normally mention unusual assessments if they provided reported research data.
Primary outcome How relevant is it to participants?	- A pragmatic trial should select an outcome with obvious importance to participants and that is measured in the same or similar way as in usual care. - Composite primary outcomes, having central adjudication and using assessments needing special training are common in explanatory trials.	- Authors usually report the primary outcome appropriately. However, sometimes the time frame is different from usual practice and this is something that could vary from country to country—an information that authors usually ignore.
Primary analysis To what extent are all data included?	- For superiority trials, pragmatic trials should conduct an intention-to-treat primary endpoint analysis. This approach, however, is also used in explanatory trials for regulated-intervention trials aiming for regulatory approval. - Other type of explanatory trials usually conduct a per protocol analysis	- Older articles do not always provide this information. - Some publications do not clearly mention the analysis performed.

that all single- and double-blind trials, and any use of placebo or sham procedures, should be considered as explanatory RCTs [2,21,22]; the use of the PRECIS-2 tool in these types of trial will only confuse readers because the total trial score obtained could hide the critical point, that is, these trials are explanatory. Furthermore, the blindness of a trial is of critical importance because potential participants are willing to join more likely in open-label trials (medicine vs no treatment) than in blind trials (medicine vs placebo) [23]. This could impact generalizability of the results by the recruitment of a biased sample of participants. Finally, we should highlight that open-label pRCTs could perfectly use blinded assessors because these latter do not alter normal clinical practice [21,22].

Regarding the generalizability of the results, this can typically only be achieved if the RCT is conducted in multiple sites that will ensure the recruitment of a heterogeneous sample of the target population with diverse demographics managed by multiple clinicians/investigators [1,2,19,24,25]. We argue that, in principle, any single-center RCT should be considered (and labeled) as explanatory, regardless the PRECIS-2 total score, providing no generalizable results. In some cases, an RCT conducted in a highly specialized center could be regarded as pragmatic because the results could be applicable to few other specialized centers [1]. However, single-center RCTs tend to provide larger intervention effects than multicenter RCTs [26−28], a key aspect when considering generalizability.

3. Three examples of the questionable use of the PRECIS-2 tool to assess the features of RCTs

Recently published studies [11,29,30] raised serious concerns on how investigators are using the PRECIS-2 tool to retrospectively assess how pragmatic/explanatory published RCTs are, even if they have correctly followed the method described by the authors of the tool—that, it should be bear in mind, was developed primarily for prospective use at the trial design stage [1]. First of all, it should be mentioned that in these three systematic reviews, the articles were the only source of information used for the assessment of the degree of pragmatism [11,29,30]. As shown in Table 1, most articles do not report many trial features needed to conduct a correct retrospective assessment with the PRECIS-2 tool. This limitation was acknowledged by the authors of two of these three analyses [11,30].

In the first one, Sepehrvand et al. [29] assessed the explanatory/pragmatic nature of 616 cardiovascular RCTs published over 2 decades (Table 2). The mean PRECIS-2 score increased significantly from 3.07 (2000) to 3.46 (2015). Following our argument described above, the results obtained in this analysis are, however, questionable; all double-blind (49% of all trials assessed), placebo-controlled (38%), and single-center trials (30%) should be considered as explanatory RCTs, regardless the PRECIS-2 total score. In addition, the FDA trial phases used in this study are appropriate for regulated-intervention trials, but not for nonregulated interventions RCTs. Regulatory agencies demand many requirements to be fulfilled in prelicensing (phase 1−3) trials that take them away from the normal clinical practice. Therefore, it would have been more reasonable to compare prelicensing (phases 2−3) vs postlicensing (phase 4) trials—and only for regulated-intervention trials.

The second study reported the trial characteristics that impacted the degree of pragmatism of 333 RCTs published in 152 nursing journals [30]. There were several types of interventions and comparators (Table 2). A score of three was given if information of a domain was missing or not applicable [18], which should have been fairly common because the only source of information was the articles assessed. The median PRECIS-2 score was 32 (somewhat pragmatic because it could take values from 9 to 45). Again, as per our point of view, this assessment is problematic because all placebo-controlled (11%), single-blind (22%), and double-blind (12%) trials should have been considered explanatory. Devos et al [30] highlighted that ‘‘blinding [the trials] resulted in no significant differences’’ in the PRECIS-2 score (scores of 32.5, 31, and 32 for single-blind, double-blind, and open-label trials, respectively, all somewhat pragmatic). However, this ‘‘remarkable’’ finding was possible because they did not consider—as we argue—that blinding is a feature that makes a trial explanatory regardless of the PRECIS-2 total score.

The third study assessed the pragmatism of 96 RCTs of advanced therapeutics in combination with methotrexate in patients with rheumatoid arthritis published in 1999−2017 [11]. In this study, only four out of the nine PRECIS-2 domains had information in all 96 trials that allowed its assessment. Only two trials provided information on recruitment and only 13 on flexibility (of intervention): adherence. Choi et al. [11] decided to score only the domains with information, so they did not provide a total score for each trial and for the whole sample of trials. They provided the mean scores of individual domains based on the actual number of trials providing information on each of them. As shown in Table 2, some were clearly explanatory, but two were pragmatic. One of these, ‘‘setting’’, scored it as pragmatic because many were multinational trials. However, the critical point to assess this domain as pragmatic is whether the trial was conducted in a setting identical to which the results will be applied [1].Therefore, authors should have known whether the sites where the trials were run were similar (e.g., specialized centers, academic centers, and hospitals) across a given trial and if these are the ones where the results will be applied. This information, however, is almost always lacking and is very difficult to check when dealing with sites located in many countries. Again, 52% of trials were placebo-controlled trials that, from our perspective, should have been considered as explanatory.

Finally, in two additional studies, 37% and 45% of single-center trials were present among RCTs of

Table 2. Recent studies in which fundamental elements of the trial design were not taken into consideration before assessing them with the PRECIS-2 tool

Study	Study description and results	Issues yielding questionable results
Sepehrvand et al. [29]	- Assessment of the explanatory/pragmatic nature of 616 cardiovascular RCTs published in six highly influential internal medicine (three) and cardiovascular (three) journals over 2 decades. - Interventions assessed: 'medicinal' (56%), 'procedure or device' (31%), and 'behavioral' or 'health system intervention' (13%). - The PRECIS-2 score could range from 1 (very explanatory) to 5 (very pragmatic) - The level of pragmatism significantly increased from a PRECIS-2 mean score of 3.07 in 2000 to 3.46 in 2015. - The increase in the PRECIS-2 score was mainly due to four domains: eligibility, setting, flexibility delivery, and primary endpoint.	- The articles were the only source of information used for the assessment of the degree of pragmatism. - 49% were double-blind trials. - 38% were placebo-controlled trials. - 30% were single-center trials. - Unjustified classification of RCTs in phase 1/2 and 3/4, which do not take into consideration that prelicensing (phases 1–3) trials have some different regulatory requirements than postlicensing (phases 4) trials. - It is not appropriate to classify nonregulated intervention trials (in this study behavioral, procedure, and health system regulated intervention) with the regulatedintervention trial phases (1 to 4). - Thirteen phase 1 trials should have been excluded; all of them were explanatory, conducted in healthy volunteers and hence not providing data of interest to the analysis.
Devos et al. [30]	- Study reporting the trial characteristics that impacted the degree of pragmatism of 333 RCTs published between 2002–2005 and 2012–2015 in 152 nursing journals. - Interventions assessed were 'therapeutics without drugs' (48%), 'therapeutic patient education' (25%), 'medication' (12%), 'care practice' (8%), and 'medical device' (7%). - The comparators were 'usual practice' (57%), 'no usual care' (16%), 'no intervention' (15%), and 'placebo' (11%). - The overall PRECIS-2 score could range from 9 (very explanatory) to 45 (very pragmatic). - The median PRECIS-2 score was 32, varying depending on the assessed interventions: from 29 for 'medication' to 33 for 'care practice' and 'therapeutic patient education'. - Trials using 'placebo' and 'no usual care' as comparators were found to have lower scores (29) than 'usual practice' (32) or 'without intervention' (34).	- The articles were the only source of information used for the assessment of the degree of pragmatism—a limitation acknowledged by the authors. - 22% were single-blind trials. - 20% were double-blind trials. - 11% were placebo-controlled trials. - 16% of trials had 'no usual care' as a comparator. This means that participants of one arm received an experimental intervention, so these trials should be considered as explanatory. - Domain: setting. Authors scored as pragmatic RCTs because they were multicenter and international. However, what matters is whether those trials were conducted in settings identical to which the results will be applied. This information is almost always impossible to check.
Choi et al. [11]	- This analysis assessed the pragmatism of 96 RCTs of biologic disease-modifying antirheumatic drugs or tofacitinib in combination with methotrexate in rheumatoid arthritis published between 1999 and 2017. - Most (94%) evaluated a specific therapeutic agent. - Most (54%) used American College of Rheumatology response criteria. - Authors only scored those domains with enough information in the articles. Authors provided the domains' overall mean scores but not the trials' overall mean score. - Eligibility, follow-up, and flexibility of intervention delivery were clearly explanatory (score around 2); conversely, setting (score: 3.6) and primary analysis (score: 4.4) were pragmatic.	- The articles were the only source of information used for the assessment of the degree of pragmatism—a limitation acknowledged by the authors. - 46% percent were placebo-controlled trials - Participants in the comparator arm were treated with active medication and placebo in 6% of trials - Domain: setting. Authors scored as pragmatic RCTs because they were multicenter and international. However, what matters is whether those trials were conducted in settings identical to which the results will be applied. This information is almost always impossible to check.

psychosocial interventions for the treatment of psychosis [14] and Chinese herbal medicine [13], respectively.

4. Future developments of the PRECIS-2 tool needed

There are two aspects that should be developed and agreed on to enhance the quality of the information provided by the PRECIS-2 tool retrospective assessments. The first deals with the domains in which inadequate or no information is provided in articles and even in protocols. In these cases, the PRECIS-2 team recommendation states that for systematic reviews, investigators should give a score of three [18]. This approach, however, will bias the total score toward three (that in PRECIS-2 score means that the domain was equally pragmatic and explanatory [1]). This is something that we are not sure will be actually useful if we aim to correctly assess the pragmatic/explanatory features of a trial.

Empirical research showed that different authors took different approaches. Thus, Devos et al. [30] gave a score of three to these domains lacking enough information. Gastaldon et al. [14] gave a score of one. These two approaches allowed providing a total score to each trial and a total score for all the trials assessed. Conversely, Choi et al. [11] decided to only score the domains from which they had enough information and gave no score to those domains with inadequate or no information; this approach provided scores to each domain and for each domain for all trials, but not a total score for each trial and, as a result, for the whole sample of trials. We believe that this latter approach is more reasonable and adjusted to what investigators know about the trials because those giving scores of one or three are assuming features that are not known at the time of the assessment. Having a total score per trial based on assumptions is not the best way to conduct a correct assessment of its pragmatic/explanatory characteristics, could lead to flawed conclusions and, we believe, should be avoided. When one or more domains are not scored in a trial because of lack of enough information, a sensible approach to obtain a total trial score would be to adjust the denominator to the number of domains that were actually scored. The issue would be to agree what would be the minimum number of domains with a score to have a total trial score acceptable in systematic reviews.

The second refers to how articles presenting reporting bias should be managed. The presence of this type of bias can only be discovered if authors cross check the information reported in the articles with that of the trial protocols and/or the registries (provided the trial was prospectively registered) [31,32]. Changes (omission, introduction, and modification) could be introduced even just before the publication of the trial results [33]. Reporting bias could relate to, among others, trial's eligibility criteria, statistical analyses, and outcome reporting. A systematic review showed that in 12% to 45% of articles, inconsistencies were observed with protocols or registries with regard to trials' selection criteria [34]. This could impact trial results by the recruitment of a different sample of participants that was initially expected at the beginning of the study. In addition, there is a high inconsistency (from 9% to 67%) in statistical analyses when comparing articles with protocols or registries [34]. Of note is that the statistical analysis model is rarely (27%) prespecified in published trial protocols [10]. Finally, and of utmost importance, in 33% of all trials, primary outcome reporting bias is present [35]. It is reasonable that logistic or protocol modifications during the conduct of a trial could shift the PRECIS-2 total score from that obtained at its planning phase [36,37]; here, however, we are addressing a very different issue: significant reporting bias could change the explanatory/pragmatic assessment of the trial. Reporting bias is of critical importance only if it is not transparent. In some trials, changes in primary outcome or analysis are perfectly acceptable provided they were included as amendments in the protocol and were reported in the article [38].

When investigators are assessing many trials with the PRECIS-2 tool—as we have referred to here [11,13,14,29,30]—, three approaches could be considered. Investigators could omit from their systematic review those trials presenting significant reporting bias (e.g., primary outcome, primary statistical analysis). Conversely, investigators could consider that their analysis should refer to what authors reported in the articles, regardless if any reporting bias is present. A third approach could be to omit in the assessment the domain (s) with significant reporting bias.

The scientific community should agree on what approach should be followed regarding the two aspects mentioned above, so the results of future studies on articles of different specialties/indications could be easily compared, something that now is not possible.

To end up with a reliable assessment, the use of PRECIS-2 tool requires training and, ideally, a discussion between raters on the trial features rather than individual scoring [39]. We strongly suggest that in retrospective analyses of published trials, the PRECIS-2 tool should only be used when additional sufficient information to that provided in articles is available. Authors should really know the details on how the trials were conducted. Because contacting trial investigators is sometimes impossible, authors should search for the available information. This latter includes but is not limited to trial protocols [40] (full documents approved by the relevant research ethics committees) and subsequent amendments or published protocols, trial registries, and, for regulated-intervention trials, EMA and FDA publicly available documents.

5. Conclusions

Citations of the PRECIS-2 tool are increasing. This useful tool to assess the pragmatic/explanatory features of RCTs could be used in retrospective analysis of published

trials if certain considerations are taken into account. The authors of the three studies discussed above [11,29,30] did not consider, as we do, that few trial design features (blinding, use of placebo, and single-center) make trials explanatory and should not be labeled as pragmatic even if the PRECIS-2 total score falls close to the pragmatic extreme. If, in addition to these design characteristics and, as happened in these three systematic reviews, authors did not have access to additional information apart from that of the articles, the assessment by means of the PRECIS-2 tool could provide misleading scores yielding questionable results.

Acknowledgments

We thank Dr Lars Hemkens (Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel, University of Basel, Basel, Switzerland) for his comments on earlier versions of this paper.Authors' contributions

R.D-R. conceived the idea and wrote the first draft of the manuscript. A.d.B. and S.K.J. provided comments and edits throughout the drafting process for important intellectual content. R.D-R., A.d.B., and S.K.J. approved the final version of the manuscript and agreed on publication of the paper.

References

[1] 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.

[2] Dal−Ré R, Janiaud P, Ioannidis JPA. Real-world evidence: how pragmatic are randomized controlled trials labeled as pragmatic? BMC Med 2018;16:49.

[3] Janiaud P, Dal-Ré R, Ioannidis JPA. Assessment of pragmatism in recently published randomized clinical trials. JAMA Intern Med 2018;178:1278−80.

[4] US National Library of Medicine. Available at ClinicalTrials.gov. Accessed May 10, 2020.

[5] Loudon K, Zwarenstein M, Sullivan FM, Donnan PT, Gágyor I, Hobbelen HJSM, et al. The PRECIS-2 tool has good interrater reliability and modest discriminant validity. J Clin Epidemiol 2017;88:113−21.

[6] Dal−Re R. Could phase 3 medicine trials be tagged as pragmatic? A case study. The Salford COPD trial. J Eval Clin Pract 2018;24:258−61.

[7] European Medicines Agency policy on publication of clinical data for medicinal products for human use. POLICY/0070. Available at https://www.ema.europa.eu/en/documents/other/european-medicines-agency-policy-publication-clinical-data-medicinal-products-human-use_en.pdf. Accessed March 21, 2019.

[8] Ladanie A, Ewald H, Kasenda B, Hemkens LG. How to use FDA drug approvals documents for evidence syntheses. BMJ 2018;362:k2815.

[9] van Rosmalen BV, Alldinger I, Cieslak KP, Wennink R, Clarke M, Ali UA, et al. Worldwide trends in volume and quality of published protocols of randomized controlled trials. PLoS One 2017;12:e0173042.

[10] Greenberg L, Jairath V, Pearse R, Kahan BC. Pre−specification of statistical analysis approaches in published clinical trial protocols was inadequate. J Clin Epidemiol 2018;101:53−60.

[11] Choi MY, Barnabe C, Barber CE, Bykerk V, Pope JE, Hazlewood GS. Pragmaticism of randomized controlled trials of biologic treatment with methotrexate in rheumatoid arthritis: a systematic review. Arthritis Care Res 2019;71:620−8.

[12] Witt CM, Manheimer E, Hammerschlag R, Ludtke R, Lao L, Tunis SR, et al. How well do randomized trials inform decision making: systematic review using comparative effectiveness research measures on acupuncture for back pain. PLoS One 2012;7:e32399.

[13] Lu L, Zhou L, Dong J, Xiang Y, Wen Z. The application of PRECIS−2 ratings in randomized controlled trials of Chinese herbal medicine. Oncotarget 2017;8:107002−10.

[14] Gastaldon C, Mosler F, Toner S, Tedeschi F, Bird VJ, Barbui C, et al. Are trials of psychological and psychosocial interventions for schizophrenia and psychosis included in the NICE guidelines pragmatic? A systematic review. PLoS One 2019;14:e0222891.

[15] Braend AM, Straand J, Klovning A. Clinical drug trials in GP & external validity. BMC Fam Pract 2017;18:113.

[16] Malmivaara A. Generalizability of findings from randomized controlled trials is limited in the leading general medical journals. J Clin Epidemiol 2019;107:36−41.

[17] Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI). Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. Lancet 1986;1:397−402.

[18] PRECIS−2. Available at Precis-2.org. Accessed May 10, 2020.

[19] Sox HC, Lewis RJ. Pragmatic trials. Practical answers to 'real world' questions. JAMA 2016;316:1205−6.

[20] Rothwell PM. External validity of randomized controlled trials: "To whom do the results of this trial apply?". Lancet 2005;365:82−93.

[21] Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ 2008;337:a2390.

[22] Welsing PM, Oude Rengerink K, Collier S, Eckert L, van Smeden M, Ciaglia A, et al. Series: pragmatic trials and real world evidence: paper 6. Outcome measures in the real world. J Clin Epidemiol 2017;90:99−107.

[23] Veerus P, Fischer K, Hemminki E, Hovi S-L, Hakama M. Effect of characteristics of women on attendance in blind and non-blind randomised trials: analysis of recruitment data from the EPHT trial. BMJ Open 2016;6:e011099.

[24] NIH collaborative. Introduction to pragmatic clinical trials. Rethinking clinical trials: a living textbook of pragmatic clinical trials. Available at https://rethinkingclinicaltrials.org/. Accessed May 10, 2020.

[25] Worsley SD, Oude Rengerink K, Irving E, Lejeune S, Mol K, Collier S, et al. Series: pragmatic trials and real world evidence: paper 2. Setting, sites, and investigator selection. J Clin Epidemiol 2017;88:14−20.

[26] Dechartres A, Boutron I, Trinquart L, Charles P, Ravaud P. Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study. Ann Intern Med 2011;155:39−51.

[27] Bafeta A, Dechartres A, Trinquart L, Yavchitz A, Boutron I, Ravaud P. Impact of single centre status on estimates of intervention effects in trials with continuous outcomes: meta-epidemiological study. BMJ 2012;344:e813.

[28] Unverzagt S, Prondzinsky R, Peinemann F. Single-center trials tend to provide larger treatment effects than multicenter trials: a systematic review. J Clin Epidemiol 2013;66:1271−80.

[29] Sepehrvand N, Alemayehu W, Das D, Gupta AK, Gouda P, Ghimire A, et al. Trends in the explanatory or pragmatic nature of cardiovascular clinical trials over 2 decades. JAMA Cardiol 2019;4:1122−8.

[30] Devos F, Foissac F, Bouazza N, Ancel PY, Tréluyer JM, Chappuy H. Study characteristics impacted the pragmatism of randomized controlled trial published in nursing: a meta-epidemiological study. J Clin Epidemiol 2019;116:18−25.

[31] Dal-Ré R, Ross JS, Marušić A. Compliance with prospective trial registration guidance remained low in high-impact journals and has implications for primary end point reporting. J Clin Epidemiol 2016;75:100−7.

[32] Gopal AD, Wallach JD, Aminawung JA, Gonsalves G, Dal-Ré R, Miller JE, et al. Adherence to the International Committee of Medical Journal Editors' (ICMJE) prospective registration policy and implications for outcome integrity: a cross-sectional analysis of trials published in high-impact specialty society journals. Trials 2018;19:448.

[33] Huić M, Marušić M, Marušić A. Completeness and changes in registered data and reporting bias of randomized controlled trials in ICMJE journals after trial registration policy. PLoS One 2011;6:e25258.

[34] Li G, Abbade LPF, Nwosu I, Jin Y, Leenus A, Maaz M, et al. A systematic review of comparisons between protocols or registrations and full reports in primary biomedical research. BMC Med Res Methodol 2018;18:9.

[35] Chen T, Li C, Qin R, Wang Y, Yu D, Dodd J, et al. Comparison of clinical trial changes in primary outcome and reported intervention effect size between trial registration and publication. JAMA Netw Open 2019;2:e197242.

[36] Johnson KE, Neta G, Dember LM, Coronado GD, Suls J, Chambers DA, et al. Use of PRECIS ratings in the national institutes of health (NIH) health care systems research collaboratory. Trials 2016;17:32.

[37] Lipman PD, Loudon K, Dluzak L, Moloney R, Messner D, Stoney CM. Framing the conversation: use of PRECIS-2 ratings to advance understanding of pragmatic trial design domains. Trials 2017;18:532.

[38] Ioannidis JPA, Caplan AL, Dal-Ré R. Outcome reporting bias in clinical trials: why monitoring matters. BMJ 2017;356:j408.

[39] Forbes G, Loudon K, Treweek S, Taylor SJC, Eldridge S. Understanding the applicability of results from primary care trials: lessons learned from applying PRECIS-2, 2017 from applying PRECIS-2. J Clin Epidemiol 2017;90:119−26.

[40] Calméjane L, Dechartres A, Ravaud P. Making protocols available with the article improved evaluation of selective outcome reporting. J Clin Epidemiol 2018;104:95−102.