CrossMark

Pragmatic randomized clinical trials: best practices and statistical guidance

Victoria Gamerman[1] · Tianxi Cai[2] · Amelie Elsäßer[3]

Abstract

Randomized clinical trials often serve the purpose of assessing the efficacy and safety of a compound. By combining real-world evidence and randomization, pragmatic randomized clinical trials (PrCTs) can be used to inform treatment effectiveness and healthcare decisions. PrCTs, referring to studies where several pragmatic elements are used (eligibility, endpoints, follow-up, etc.), pose unique challenges (Loudon et al. in BMJ 350:h2147, 2015). From a literature review, we propose a definition of PrCT and discuss strategies to overcome some PrCT challenges. Use of alternative data collection approaches may lead to uncertainties, and absence of blinding could potentially lead to non-random missing data at study endpoints such that randomization is no longer protected by an intent to treat. Therefore, more complex randomization strategies may be needed to minimize bias. Additional data sources could be used to synthesize information and create a more accurate endpoint definition, which may require tools such as natural language processing. The statistician must become familiar with the challenges and strengths of PrCTs, ranging from design to analysis to interpretation, in order to transform data into evidence (Califf in Clin Trials 13:471–477, 2016).

1 Introduction

Classical randomized clinical trials (RCTs), especially Phase III trials, are planned with the intent to show efficacy and safety of a new compound. Through the use of a large list of inclusion and exclusion criteria, the enrolled patient population involved in these trials is usually very homogenous, which ensures the rigor necessary to eliminate factors that may

✉ Victoria Gamerman
   victoria.gamerman@boehringer-ingelheim.com

1   Boehringer Ingelheim Pharmaceuticals, Inc., 900 Ridgebury Rd, Ridgefield, CT 06877, USA

2   Department of Biostatistics, Harvard T.H. Chan School of Public Health, 665 Huntington Avenue, Boston, MA 02115, USA

3   Boehringer Ingelheim Pharma GmbH & Co. KG, Binger Str. 173, 55216 Ingelheim am Rhein, Germany

introduce bias. The sites contributing are mainly specialized clinical research study sites. Often, these sites have been running trials for several years, have staff with extensive clinical trial experience, and are familiar with the processes involved in running such studies, from patient recruitment, informed consent, ensuring regular visit schedules, up to correct data collection [with standard electronic case report forms (eCRFs), or even via personal devices such as ambulatory blood pressure measurement monitors].

As a result, classical RCTs involve a special population and are conducted in a widely artificial environment, which can differ from how the majority of patients are treated in practice. Pragmatic randomized clinical trials (PrCTs) are conducted to answer the important question of how a treatment works in a "real-world scenario," a heterogeneous "real-world population," in short to estimate the treatment's effectiveness. The concept of PrCTs is not new, as pragmatic trial approaches were discussed as far back as the 1960s (Schwartz and Lellouch 1967).

Real-world data are often equated with observational trial data, but "real-world evidence and randomization are two fully compatible concepts" as highlighted by Sherman and colleagues from the Food and Drug Administration (FDA) in their 2016 publication (Sherman et al. 2016). Randomization is a necessary concept as we will highlight in this manuscript (section design aspects). The importance of randomization and the role of the statistician was highlighted by the former FDA Commissioner Califf in 2016: "Statisticians can also perform a valuable service by continually reminding people about what a powerful tool randomization is" (Califf 2016).

Here we discuss PrCT definitions in the literature, propose a broad definition anchored on concepts used by others (including Hotopf 2002; MacPherson 2004; Zuidgeest et al. 2017; the CONSORT statement updated for pragmatic studies (Zwarenstein et al. 2008); and the PRECIS tool (Thorpe et al. 2009)], and highlight several statistical challenges with potential solutions.

2 Literature review

To assess the environment of PrCTs, a search was conducted in the ClinicalTrials.gov database. The search results from ClinicalTrials.gov are presented below and at the 2018 International Conference on Health Policy Statistics organized by the Health Policy Statistics Section of the American Statistical Association. A set of four search terms was used: (1) "pragmatic" and "randomized"; (2) "pragmatic" and "randomised"; (3) "real world" and "randomized"; and (4) "real world" and "randomised."

2.1 ClinicalTrials.gov database search

In September 2017, a search of trials entered in the ClinicalTrials.gov database was conducted. Our objective was to determine how readily identifiable PrCTs were in the database. Four search terms were used, and 732 unique trials were identified after excluding duplicates. Table 1 shows the search terms and count obtained for each search result, which was downloaded to a spreadsheet from the database.

Several additional exclusions were made to reduce the list, thereby making it feasible to review. To narrow down the list of identified studies, studies that were not listed as interventional (n = 40) or randomized (n = 29) were excluded. At the time of this review, there was an interest in gaining insight into interventional, randomized,

Table 1 Search terms used to pull trials from the ClinicalTrials.gov database (September 2017)

Search term	Count of results
"pragmatic" and "randomized"	450
"pragmatic" and "randomised"	94
"real world" and "randomized"	271
"real world" and "randomised"	15

and industry-sponsored studies to see if other pharmaceutical companies were conducting studies in addition to the popular Salford Lung Study example (New et al. 2014). By focusing on industry-sponsored studies (excluding another 580 studies), 83 unique entries remained on the list (Fig. 1). Upon reviewing the simple titles extracted from the database search of industry-sponsored entries, we found that only 20 simple title reviews indicated that they were potentially PrCTs. Of these 20 trials, most were listed as Phase IV and were in primary care therapeutic areas (mostly either in metabolism or respiratory). This finding of just 20 studies from the list of 83 suggests that quickly identifying PrCTs based on a simple title review alone is not sufficient to understand the landscape of PrCTs being conducted. Overall, identification of pragmatic trials is subjective and requires further in-depth review. Table 2 shows the list of the final 20 studies identified. Excluding the 83 industry-sponsored studies among the 663 randomized and interventional studies, a random sample of the remaining 580 studies that were not industry sponsored was taken. The random sample of 54 studies was pulled and a review of the detailed title was conducted. Here, it was found that 50 of the studies were PrCTs. This confirms that a review of the simple title alone is not sufficient and indicates that a review of the detailed official title is a better source of information on pragmatic characteristics of posted trials. Moreover, a larger proportion of the identified non-industry sponsored trials were PrCTs compared to the industry sponsored trials, possibly because trials with pragmatic features may be even more challenging to accomplish in a pre-approval setting for the industry sponsored trials.

This search revealed that pragmatic trials are neither intuitive nor easy to identify through a database search of ClinicalTrials.gov if relevant terms such as "pragmatic" or "real world" are not used in the title. Inclusion of real-world studies in ClinicalTrials.gov was a topic of discussion at the joint summit of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and International Society for
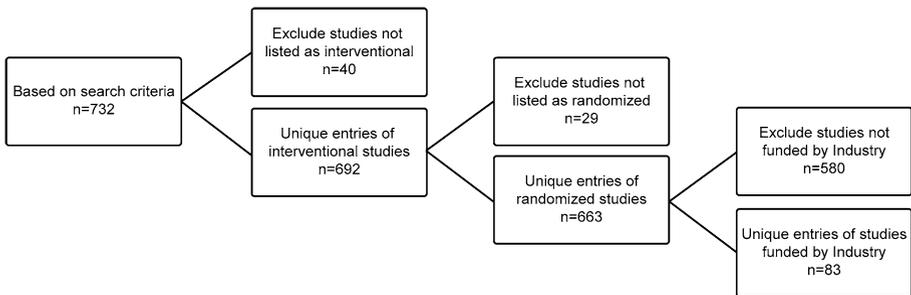

Fig. 1 Narrowing the search results from ClinicalTrials.gov to identify PrCTs in the database as of September 2017. *PrCT* pragmatic randomized clinical trial

Table 2 List of PrCTs identified for analysis

Title	NCT number
A Cluster-randomized, Pragmatic Trial of Hemodialysis Session Duration	NCT02019225
A Pragmatic Trial to Evaluate the Comparative Effectiveness Between Dapagliflozin and Standard of Care in Type 2 Diabetes Patients (DECIDE Study)	NCT02616666
Assessment In a Real World Setting of the Effect of Inhaled Steroid-based Triple Therapy Versus the Combination of Tiotropium and Olodaterol on Reducing Chronic Obstructive Pulmonary Disease (COPD) Exacerbations [AIRWISE]	NCT03265145
Pragmatic RCT Comparing Aripiprazole, Olanzapine and Haloperidol in the Treatment of Schizophrenia	NCT01052389
To Evaluate Effectiveness of Aclidinium Bromide/Formoterol Fumarate Dihydrate in Chronic Obstructive Pulmonary Disease	NCT03181880
Effectiveness Trial for Evaluating IAHA for PFPS	NCT01811654
A "Real World" Trial to Determine Efficacy and Health Outcomes of Toujeo (ACHIEVE CONTROL REAL LIFE STUDY PROGRAM)	NCT02451137
A "Real World" Trial to Determine Efficacy and Health Outcomes of Toujeo Compared to "Standard of Care" Basal Insulin in Patients Already Using Basal Insulin	NCT02967211
A "Real World" Trial to Determine Efficacy and Health Outcomes of Toujeo Compared to "Standard of Care" Basal Insulins in Insulin Naïve Patients Initiating Insulin	NCT02967224
Pragmatic Clinical Trial for a More Effective Concise and Less Toxic MDR-TB Treatment Regimen(s)	NCT02589782
Reducing Clinical Inertia in Hypertension Treatment: A Pragmatic Trial	NCT01145391
A Pragmatic Trial of Corticosteroid Optimisation in Severe Asthma	NCT02717689
Effectiveness and Cost-effectiveness of Sacral Neuromodulation in Patients With Idiopathic Slow-transit Constipation	NCT02961582
Effectiveness of Two Hyaluronic Acids in Osteoarthritis of the Knee	NCT01543737
A Comparative Effectiveness Study of Major Glycemia-lowering Medications for Treatment of Type 2 Diabetes	NCT01794143
Real-World AR101 Market-Supporting Experience Study in Peanut-Allergic Children	NCT03126227
Affordability and Real-world Antiplatelet Treatment Effectiveness After Myocardial Infarction Study	NCT02406677
Modeling and Predicting Real World Behavior Using Mobile Sensor Data on Patients With Major Depressive Disorder	NCT02499094
Comparison of BuMA eG Based BioDegradable Polymer Stent With EXCEL Biodegradable Polymer Sirolimus-eluting Stent in "Real-World" Practice	NCT02017275
Improving Asthma Control in the Real World: A Systematic Approach to Improving Dulera Adherence	NCT02045875

PrCT pragmatic randomized clinical trial

Pharmacoepidemiology (ISPE) on "Real-World Evidence in Health Care Decision Making," Washington DC, October 20, 2017.

In a recent publication, Dal-Ré et al. (2018) present a search for "pragmatic" and "naturalistic" trials on PubMed. They suspect that many trials labeled as pragmatic are not necessarily very pragmatic. Dal-Ré et al. therefore recommend a PRECIS-2 assessment for all pragmatic trials to be provided by the trial investigators for publication purposes.

To make it easier to identify pragmatic trials (e.g., in database searches), first, a clear definition is needed. In the following section, different trial definitions from the literature are presented and a proposal for a PrCT definition is given.

2.2 Pragmatic randomized trial definitions in literature

There are several aspects that are discussed when it comes to the definition of PrCTs. The definitions often are broad and no single, clear definition is used consistently. These can be categorized as follows:

2.2.1 Population

The population is heterogeneous with patients being similar to the patients that would be expected to be treated with the new therapy in practice. Enrolling a patient population reflective of the real-world population is one of the key elements of a PrCT study design. Unlike a controlled clinical trial, where the inclusion and exclusion criteria help to define the perfect patient to demonstrate treatment efficacy and safety, in the PrCT setting, the aim is to identify a heterogeneous patient population to optimize the generalizability of the trial results.

During the design phase of the study, consideration should be given to how this cohort will be identified and how to subsequently retain their enrollment in the pragmatic trial. One concept of consecutive enrollment describes consecutively enrolling all patients who are eligible to receive the intervention as part of their routine care within a clinic. Bias can also be introduced due to low enrollment and loss to patient follow-up. Such bias can undermine the external validity of the study, making the results less generalizable, which is the major reason for considering a real-world population in this analysis.

Generally, it is recommended to have minimal inclusion and exclusion criteria to identify a heterogeneous patient population for PrCT. Discussion on patients who are enrolled in the PrCT while also enrolled in other studies should be conducted as part of defining the real-world population eligible for the trial (Rengerink et al. 2017).

2.2.2 Setting

The PrCT setting should be representative of and similar to the setting in which patients are treated in the real world. "The staff should have typical experience working with these patients, although research experience is not a requisite. Recruitment may be slow when working at sites that perform routine clinical care as opposed to research sites (Rengerink et al. 2017). Therefore, maximizing the similarity to routine clinical care could positively influence the sites' ability to enroll patients and subsequent patient participation. Engaging the sites to best understand their routine care during the protocol development phase may also alleviate future challenges (Rengerink et al. 2017). Upon defining the routine or standard-of-care for a real-world patient population, consideration should be given to the diversity of the sites and the patient population that may be enrolled at these sites. Certain site characteristics will also influence the successful outcome of a PrCT, such as the availability of staff at a clinical practice, experience in conducting research, and availability of patients. As these sites may not regularly conduct research studies, additional support in terms of tools and consistency with routine practices may be necessary. Involving sites in the protocol development may also improve long-term engagement (Worsley et al. 2017).

2.2.3 Data collection and outcomes

Data collection can be done via standard eCRFs, via electronic health records (EHRs), or personal devices. The number of endpoints can range between only very few but meaningful endpoints and a multitude of endpoints, e.g., those collected via an EHR. Here, EHR refers to an electronic version of a patient's paper medical record containing information on the patient's overall health.

To best inform healthcare decisions, outcomes captured in a PrCT should reflect the information needed to make an informed decision by patients and physicians during routine care. Therefore, not all the endpoints of a controlled clinical trial are of the same importance for a PrCT. One such consideration is that capturing an outcome in the pragmatic setting should be in line with routine care and easily assessable (e.g., morbidity, mortality, and resource utilization). The outcome should also be patient relevant so that informed decisions on the treatment can be made by both doctors and patients cooperatively. Often PrCTs are open-label; therefore, hard, objective endpoints should be used and captured in a timely and precise manner to minimize bias. Blinding and standardization will be important to minimize bias from data collection in an open-label setting as well (Welsing et al. 2017).

Measuring the outcome will be an important topic to discuss during study design to ensure that the relevant outcomes are feasible to capture. Data may also be collected passively by extracting information directly from the EHR.

2.2.4 Patient-relevant design

Trials are randomized but often open-label to represent the real-world setting in which the patient is aware of the treatment received. To best inform healthcare decisions, outcomes captured in a PrCT should reflect relevant outcomes. The outcomes should be specific for the population of interest and should help compare the different treatments. A patient-relevant comparison of intervention versus control arms should also help in making informed treatment decisions for future patients. This can be achieved by using an appropriate comparison arm in the study design. One needs to make sure that the outcomes can be captured during both routine clinical care and the trial period (Zuidgeest et al. 2017). Special trial designs can be considered, such as platform trials to embed PrCTs within a learning healthcare system, the use of complex interventions, or cluster randomization (see Sect. 3.2).

Overall, many trials have pragmatic as well as explanatory elements included. There is a continuum between these two approaches (NIH collaboratory n.d.; Loudon et al. 2015). Very few trials are fully pragmatic in every design aspect (Ford and Norrie 2016).

2.3 Pragmatic trial definition proposal

With the diversity and variety in definitions in the literature, we propose the following pragmatic randomized trial definition.

A PrCT is an RCT with *four key pragmatic design elements* (*see also Fig.* 2):

Fig. 2 Key design elements of a pragmatic randomized clinical trial

Real-world population
Real-world setting
Appropriate comparison arm
Relevant outcome
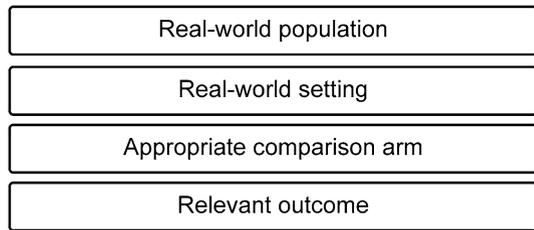
1. The trial should enroll a *real-world population*, i.e., a population close to the patient population that would receive the treatment in practice, making this PrCT population more generalizable than the one seen in standard RCTs.
2. The trial should always be conducted in a *real-world setting,* i.e., practitioners as investigators and community sites rather than research study sites, although this depends on the indication of interest.
3. The trial should include an *appropriate comparison arm* depending on the question of interest. Often, it might be an active-control arm instead of a placebo arm.
4. The trial should capture the *relevant outcomes* to inform optimal healthcare treatment decisions at the end. The outcomes should be relevant and the number of endpoints should be limited to the key endpoints to preserve a simple trial approach. The data can be captured in different ways, not solely through eCRFs. Data from EHRs can provide helpful information without large extra burden for the investigator. Personal devices can provide real-time data processing and have the asset of being close to the patient.

3 Statistical design considerations

There are important design and analysis aspects that should be considered when planning a pragmatic trial. We first detail whether and how randomization and blinding can be employed in PrCT to minimize bias. We will also detail the additional challenges arising from endpoint capturing in the pragmatic setting. Analytical strategies to overcome missing data and imperfect endpoint are discussed at the end of Sect. 3.

3.1 Randomization

In clinical trials, randomization helps to ensure that the patients in the different treatment arms are comparable with regard to their baseline characteristics. Any difference in baseline characteristics seen between the treatment arms is solely due to chance in an RCT setting. In this way, the observed between-arm difference can be interpreted as the causal effect of the treatment. Randomization removes relevant differences in the measured parameters as well as even more importantly, unmeasured factors at baseline.

As in standard RCT, randomization in PrCTs can either be performed on a patient level, which is standard in clinical trials in general, as well as on a cluster level, e.g., on a site, clinic, or doctor level. Another new design for PrCTs is the cohort multiple control randomized studies (Relton et al. 2010). In these designs, a cohort of patients is followed over time, a subgroup (based on exclusion criteria) of the cohort is selected and then a random group of patients within the specific subgroup is invited to participate in a trial and

to receive a new treatment. The patients in the random group are then compared with the patients from the subgroup who were not selected and were treated (e.g., with standard of care). This process can be repeated over time for conducting different trials within changing subgroups within the cohort. These trials offer the opportunity for special informed consent (IC) concepts, e.g., one IC for all the patients in the cohort and one for all the patients in the trial. Another opportunity of this type of design is that it allows brand new therapies becoming available to be investigated straight away in time in a representative real-world population subgroup in case the general trial is set up already and the cohort is already followed up.

3.2 Blinding

Blinding is a useful strategy that can minimize the impact of subjective factors on the trial outcomes. Blinding can avoid subjective bias in patients' participation and reporting of the outcome. This is particularly important for trials with endpoints that are subjective. Similarly, during the treatment period of a randomized trial, information on patient assignment to treatment groups might influence the care administered by health care providers (HCPs). Therefore, blinding of HCPs to treatment assignments would theoretically avoid bias in clinical care of patients and would not influence the outcome of one group more than the other (Viera and Bangdiwala 2007). Blinding can lead to an overall higher internal validity compared to similar trials without blinding. Blinding, together with randomization, can ensure that, the observed treatment effect is caused by the therapy (Fransen et al. 2007); this leads to more reliable results and to a higher credibility (Godwin et al. 2003).

However, blinding is not always feasible (Day and Altman 2000). In some situations, blinding cannot be performed for ethical reasons, e.g., when a surgery is compared with a drug treatment, the patients randomized to the drug treatment arm would need to get sham surgery to keep the blind, which raises ethical concerns and might therefore not be acceptable. Incorporating blinding is also particularly challenging for PrCTs that target broad patient populations and are often embedded in standard healthcare settings. Patients may prefer non-blinded trials over blinded trials as seen in the Estonian Postmenopausal Hormone Therapy (EPHT) trial (Veerus et al. 2016). The EPHT trial evaluated hormone replacement therapy in postmenopausal women in two parallel subtrials, one blinded and one non-blinded. Women were randomized to one of the subtrials. The result was that "women randomized to the non-blinded subtrial were more willing to join than women in the blinded subtrial," and "women with higher education were differentially more willing to join the non-blinded trial." Thus, enforcing blinding may not be desirable if it results in a patient population that is less general than seen in practice (Godwin et al. 2003; Fransen et al. 2007; Macpherson 2004). This is particularly important for PrCTs where generalizability is a key issue. When a PrCT is embedded in a standard clinical care setting, it may also be infeasible to blind the clinician managing the care of the patient.

When the patient or investigator cannot be blinded, the following strategies can be used to minimize the potential subjective bias:

1. If possible, only hard endpoints (no subjective endpoints!) should be used for primary and key secondary endpoints.
2. In situations where event- or outcome-based endpoints are used, adjudication by blinded medical experts is recommended.

3. If possible, the statistician/data analyst should be blinded during the trial conduct, during the statistical analysis plan development, and beyond for as long as possible. If any interim analyses are performed, these should be conducted in a blinded fashion by the statistician.

Cluster randomization can be discussed and can help if blinding of individuals is not reasonable (Godwin et al. 2003; Fransen et al. 2007). Thought should be given to any potential contamination of study data in the unblinded PrCT setting, which may occur during the study conduct when, for example, patients in the control or usual care arm actively seek out the intervention being investigated. The reality of this contamination will depend on the feasibility of the patients being able to get access to the intervention since it may be an expensive or a special procedure, with limited access to a broader audience.

3.3 Capture of endpoint

A major challenge arising from PrCTs embedded in a standard clinical care setting is that the clinical endpoint may not be manually recorded by study investigators. Under this setting, the investigators rely on the healthcare system to capture the endpoint information. As such, the endpoint of interest may not be precisely or fully observed. The endpoint information would be missing if a patient does not return to the healthcare system during the study visit times. Even among the patients who have information recorded during planned visit times, outcome information recorded in the EHR often cannot be automatically annotated precisely. There are two reasons for this: first, the diagnostic billing codes are often imprecise with varying degrees of accuracy in reflecting the true status of the clinical condition; second, the outcome information may be documented in the narrative notes. Natural language processing (NLP), while powerful, is an imperfect tool for extracting information related to a condition of interest (Nadkarni et al. 2011). Precise information on the outcomes of interest still requires labor-intensive human annotation (Pathak et al. 2013). The naïve use of imprecise and partially missing endpoint information may lead to a bias in the treatment effect assessment.

3.4 Data augmentation

To minimize the impact of missing data due to lack of information in the EHR, additional tools such as mobile devices or online portals can be used to capture the endpoint as patient-reported outcomes. To ensure the validity of the PrCT findings, one plausible approach is to have a small number of local sites to additionally collect endpoint data using traditional mechanisms as in a standard RCT. Such data can be used as validation data for assessing and correcting for potential biases induced by missing or imprecise data on the endpoint.

To overcome the challenge of imprecise endpoint information extraction and the infeasibility of large-scale manual chart review, one may employ machine learning algorithms to predict the outcome using partially labeled data (Liao et al. 2015). Specifically, the study investigators may manually annotate a small number of gold standard labels on the outcome and extract predictive features such as the billing codes and NLP mentions of relevant clinical terms. Although the precise information on the outcome is only available for the labeled set, the EHR features can be easily extracted for all patients. Then a machine learning algorithm can be developed to predict the outcomes for the unlabeled patients using

their EHR features. For binary outcomes, such algorithms can also provide a probabilistic annotation to reflect uncertainty in the prediction, which can potentially improve the power of subsequent association studies (Sinnott et al. 2014). Despite the recent advancement in EHR phenotyping, phenotyping algorithms cannot perfectly classify the outcome of interest in most cases.

3.5 Statistical analyses

Both the missingness and the imprecision pose challenges to the analysis of PrCT data. If the endpoint of interest can be ascertained precisely for all patients, one may perform the intent-to-treat (ITT) analysis for PrCTs to infer causal effects. However, due to the potential lack of blinding, patients may have differential patterns of missingness in their endpoint across different treatment groups. As such, standard complete case analysis may result in biased treatment effect estimates despite the randomization. To correct for bias arising from confounding due to differentially missing data, causal inference tools such as propensity score adjustment, marginal structural modeling, and doubly robust methods (Rosenbaum and Rubin 1983; Robins et al. 2000; Bang and Robins 2005) can be used. Because of differential missingness, treatment arms may not exhibit balanced distribution of patients, leading to potential bias and inaccurate results. Use of propensity scores allows researchers to determine the probability of treatment administration based on observable covariates and helps restore balance (Malla et al. 2018). Hence, if missing values are meaningful for analysis, a statistically valid analysis performed under a primary set of missing-data assumptions and a sensitivity analysis for assessing robustness of inferences to missing-data assumptions should be used to draw inferences from incomplete data (Little et al. 2012). Additional challenges arise from the imprecision in the endpoint derived from phenotyping algorithms. This can be viewed as a mismeasured outcome, and inference tools that incorporate measurement errors in the outcomes can be employed to correct for the potential bias (Sinnott et al. 2014; Caroll et al. 2006).

4 Discussion

Differing from classical RCTs, pragmatic trials answer the important question of the effectiveness of a therapy in the "real world," rather than efficacy in a pre-specified patient population. One of the main advantages in conducting PrCTs is the importance of the use of randomization, which cannot be implemented in real-world observational studies. In conducting a literature review, it was clear that diverse definitions of PrCTs are currently being used. Further, we conducted searches in the ClinicalTrials.gov database (conducted in September 2017) and identified several hundred trials based on the search terms used, but not all these trials can be referenced as pragmatic trials, as highlighted by Dal-Ré et al. (2018). As a result, a simple definition for PrCTs was proposed with four key design elements: enrolling a *real-world population*, conducted in a *real-world setting*, capturing *relevant outcomes* important to inform optimal healthcare treatment decisions, and using an *appropriate comparison arm*, which may not always be a placebo treatment. A study meeting this definition would not necessarily be the best design to address safety and efficacy, but the results would be able to offer an insight into the real-world effectiveness of an intervention.

While many studies have design elements spanning the exploratory-pragmatic continuum, this broad definition should be considered as a minimum for a trial to be labeled pragmatic.

In general, it is hard to identify pragmatic trials in these database searches using search terms, and an in-depth review of the information posted on ClinicalTrials.gov would be needed to classify them as pragmatic trials. Therefore, the suggestion that trials in databases should also include a score like the PRECIS-2 score would be the first step toward improving judgments on the pragmatism of a clinical trial. In addition, researchers could harmonize the definition and consideration of PrCTs by making the entry and search in the ClinicalTrials.gov database more intuitive and straightforward for these real-world studies. If there were a field or other way to self-identify trials as PrCTs, the National Institutes of Health (NIH) would have the opportunity to spell out a definition as is done for other fields in the database. This topic of including real-world studies in the database was discussed at the ISPOR/ISPE joint summit on "Real-World Evidence in Health Care Decision Making" in October 2017.

Important design issues for PrCT include randomization and blinding. Randomization could be considered in a PrCT at the individual or cluster level or even in a cohort multiple control randomized study design. Performing blinded PrCT may be particularly challenging due to its feasibility and potential impact on the participating patient population. When blinding is employed, one needs to also consider operational complexities, including interactive response technology and drug dispensation, which may make the trials less pragmatic and more expensive to carry out. In cases where blinding was not feasible, special care needs to be given to the design in order to minimize potential subjective bias. Although Phase III RCTs are often global studies, additional consideration should be given to the generalizability of a PrCT conducted within a single country or region, which may be necessary due to local regulations (for data sharing) or EHR system limitations.

Analysis of PrCT could be more challenging compared to the standard RCTs. The ITT analysis may no longer be appropriate for PrCT when systematic differences between different treatment arms, such as patterns of missingness in the outcome, exist. Using the EHR to capture study endpoints can be limiting if its precise information cannot be extracted accurately. Supervised or semi-supervised machine learning methods can be used to train an algorithm to predict the outcome, but regardless of the approach used, these algorithms are likely to yield misclassification and hence measurement error in the outcomes. Statistical methods for drawing causal inference and accounting for outcome measurement error can be used to adjust for factors that contribute to missingness and measurement error.

The future can be very bright for the use of PrCTs to better understand the effectiveness of treatments in the real world, although this would not be the ideal setting to assess efficacy. To help facilitate the dialogue, a simple definition for PrCTs is proposed. Significant consideration during the design phase should be given to elements of the study that will impact how the data collected can be affected and analyzed.

Compliance with ethical standards

Conflict of interest VG and AE are employees of Boehringer Ingelheim Pharmaceuticals, Inc. TC has no conflict of interest to declare.

Disclosures The author(s) meet criteria for authorship as recommended by the International Committee of Medical Journal Editors (ICMJE). The authors received compensation (i.e., salary) as employees of

References

Bang, H., Robins, J.M.: Doubly robust estimation in missing data and causal inference models. Biometrics **61**, 962–973 (2005)

Califf, R.M.: Pragmatic clinical trials: emerging challenges and new roles for statisticians. Clin. Trials. **13**, 471–477 (2016)

Caroll, R.J., Rupert, D., Stefanski, L.A., Crainiceanu, C.M.: Measurement Error in Nonlinear Models: A Modern Perspective. Chapman & Hall/CRC, Taylor and Francis Group, New York (2006)

Dal-Ré, R., Janiaud, P., Ioannidis, J.P.: Real-world evidence: how pragmatic are randomized controlled trials labeled as pragmatic? BMC Med. **16**, 49 (2018)

Day, S.J., Altman, D.G.: Statistics notes: blinding in clinical trials and other studies. BMJ **321**, 504 (2000)

Ford, I., Norrie, J.: Pragmatic trials. N. Engl. J. Med. **375**, 454–463 (2016)

Fransen, G.A., van Marrewijk, C.J., Mujakovic, S., Muris, J.W., Laheij, R.J., Numans, M.E., de Wit, N.J., Samsom, M., Jansen, J.B., Knottnerus, J.A.: Pragmatic trials in primary care. Methodological challenges and solutions demonstrated by the DIAMOND-study. BMC Med. Res. Methodol. **7**, 16 (2007)

Godwin, M., Ruhland, L., Casson, I., MacDonald, S., Delva, D., Birtwhistle, R., Lam, M., Seguin, R.: Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. BMC Med. Res. Methodol. **3**, 28 (2003)

Hotopf, M.: The pragmatic randomised controlled trial. Adv. Psychiatr. Treat. **8**, 326–333 (2002)

Liao, K.P., Cai, T., Savova, G.K., Murphy, S.N., Karlson, E.W., Ananthakrishnan, A.N., Gainer, V.S., Shaw, S.Y., Xia, Z., Szolovits, P., Churchill, S., Kohane, I.: Development of phenotype algorithms using electronic medical records and incorporating natural language processing. BMJ **350**, h1885 (2015)

Little, R.J., D'Agostino, R., Cohen, M.L., Dickersin, K., Emerson, S.S., Farrar, J.T., Frangakis, C., Hogan, J.W., Molenberghs, G., Murphy, S.A., Neaton, J.D., Rotnitzky, A., Scharfstein, D., Shih, W.J., Siegel, J.P., Stern, H.: The prevention and treatment of missing data in clinical trials. N. Engl. J. Med. **367**, 1355–1360 (2012)

Loudon, K., Treweek, S., Sullivan, F., Donnan, P., Thorpe, K.E., Zwarenstein, M.: The PRECIS-2 tool: designing trials that are fit for purpose. BMJ **350**, h2147 (2015)

Macpherson, H.: Pragmatic clinical trials. Complement. Ther. Med. **12**, 136–140 (2004)

Malla, L., Perera-Salazar, R., McFadden, E., Ogero, M., Stepniewska, K., English, M.: Handling missing data in propensity score estimation in comparative effectiveness evaluations: a systematic review. J. Comp. Eff. Res. **7**, 271–279 (2018)

Nadkarni, P.M., Ohno-Machado, L., Chapman, W.W.: Natural language processing: an introduction. J. Am. Med. Inform. Assoc. **18**, 544–551 (2011)

New, J.P., Bakerly, N.D., Leather, D., Woodcock, A.: Obtaining real-world evidence: the Salford Lung Study. Thorax **69**, 1152–1154 (2014)

NIH Collaboratory. Introduction to pragmatic clinical trials (n.d.). https://www.nihcollaboratory.org/Products/Introduction%20to%20pragmatic%20clinical%20trials.pdf. Accessed 25 April 2018

Pathak, J., Kho, A.N., Denny, J.C.: Electronic health records-driven phenotyping: challenges, recent advances, and perspectives. J. Am. Med. Inform. Assoc. **20**, e206–e211 (2013)

Precis-2. https://www.precis-2.org/Trials. Accessed 25 April 2018

Relton, C., Torgerson, D., O'Cathain, A., Nicholl, J.: Rethinking pragmatic randomised controlled trials: introducing the "cohort multiple randomised controlled trial" design. BMJ **340**, c1066 (2010)

Rengerink, K.O., Kalkman, S., Collier, S., Ciaglia, A., Worsley, S.D., Lightbourne, A., Eckert, L., Groenwold, R.H., Grobbee, D.E., Irving, E.A.: Series: pragmatic trials and real world evidence: paper 3. Patient selection challenges and consequences. J. Clin. Epidemiol. **89**, 173–180 (2017)

Robins, J.M., Hernan, M.A., Brumback, B.: Marginal structural models and causal inference in epidemiology. Epidemiology **11**, 550–560 (2000)

Rosenbaum, P.R., Rubin, D.B.: The central role of the propensity score in observational studies for causal effects. Biometrika **70**, 41–55 (1983)

Schwartz, D., Lellouch, J.: Explanatory and pragmatic attitudes in therapeutical trials. J. Chronic. Dis. **20**, 637–648 (1967)

Sherman, R.E., Anderson, S.A., Dal Pan, G.J., Gray, G.W., Gross, T., Hunter, N.L., LaVange, L., Marinac-Dabic, D., Marks, P.W., Robb, M.A., Shuren, J., Temple, R., Woodcock, J., Yue, L.Q., Califf, R.M.: Real-world evidence—what is it and what can it tell us? N. Engl. J. Med. **375**, 2293–2297 (2016)

Sinnott, J.A., Dai, W., Liao, K.P., Shaw, S.Y., Ananthakrishnan, A.N., Gainer, V.S., Karlson, E.W., Churchill, S., Szolovits, P., Murphy, S., Kohane, I., Plenge, R., Cai, T.: Improving the power of genetic association tests with imperfect phenotype derived from electronic medical records. Hum. Genet. **133**, 1369–1382 (2014)

Thorpe, K.E., Zwarenstein, M., Oxman, A.D., Treweek, S., Furberg, C.D., Altman, D.G., Tunis, S., Bergel, E., Harvey, I., Magid, D.J., Chalkidou, K.: A pragmatic–explanatory continuum indicator summary (PRECIS): a tool to help trial designers. J. Clin. Epidemiol. **62**(5), 464–475 (2009)

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 **6**, e011099 (2016)

Viera, A.J., Bangdiwala, S.I.: Eliminating bias in randomized controlled trials: importance of allocation concealment and masking. Fam. Med. **39**, 132–137 (2007)

Welsing, P.M., Rengerink, K.O., Collier, S., Eckert, L., van Smeden, M., Ciaglia, A., Nachbaur, G., Trelle, S., Taylor, A.J., Egger, M., Goetz, I.: Series: pragmatic trials and real world evidence: paper 6. Outcome measures in the real world. J. Clin. Epidemiol. **90**, 99–107 (2017)

Worsley, S.D., Rengerink, K.O., Irving, E., Lejeune, S., Mol, K., Collier, S., Groenwold, R.H., Enters-Weijnen, C., Egger, M., Rhodes, T.: Series: pragmatic trials and real world evidence: paper 2. Setting, sites, and investigator selection. J. Clin. Epidemiol. **88**, 14–20 (2017)

Zuidgeest, M.G., Welsing, P.M., van Thiel, G.J., Ciaglia, A., Alfonso-Cristancho, R., Eckert, L., Eijkemans, M.J., Egger, M.: Series: pragmatic trials and real world evidence: paper 5. Usual care and real life comparators. J. Clin. Epidemiol. **90**, 92–98 (2017)

Zwarenstein, M., Treweek, S., Gagnier, J.J., Altman, D.G., Tunis, S., Haynes, B., Oxman, A.D., Moher, D.: Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ **337**, a2390 (2008)