Manchester Metropolitan University

e-space

Manchester Metropolitan University's Research Repository

**Please cite the Published Version**

Palin, Victoria, Van Staa, Tjeerd P, Steels, Stephanie, Troxel, Andrea B, Groenwold, Rolf HH, MacDonald, Tom M, Torgerson, David, Faries, Douglas, Mancini, Pierre, Ouwens, Mario, Frith, Lucy J, Tsirtsonis, Kate, MacLennan, Graham and Nordon, Clementine (2022) A first step towards best practice recommendations for the design and statistical analyses of pragmatic clinical trials: a modified Delphi approach. British Journal of Clinical Pharmacology, 88 (12). pp. 5183-5201. ISSN 0306-5251

**DOI:** https://doi.org/10.1111/bcp.15441

**Publisher:** Wiley

**Version:** Accepted Version

**Downloaded from:** https://e-space.mmu.ac.uk/630100/

**Usage rights:** © In Copyright

**Additional Information:** This is the peer reviewed version of the following article: Palin, V, Van Staa, TP, Steels, S, et al. A first step towards best practice recommendations for the design and statistical analyses of pragmatic clinical trials: A modified Delphi approach. Br J Clin Pharmacol. 2022, which has been published in final form at https://doi.org/10.1111/bcp.15441. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. This article may not be enhanced, enriched or otherwise transformed into a derivative work, without express permission from Wiley or by statutory rights under applicable legislation. Copyright notices must not be removed, obscured or modified. The article must be linked to Wiley's version of record on Wiley Online Library and any embedding, framing or otherwise making available the article or pages thereof by third parties from platforms, services and websites other than Wiley Online Library must be prohibited.

**Data Access Statement:** No human subject participated in the present study. The data generated and analysed during the current study are available in the Google Drive repository without any restriction: https://drive.google.com/drive/u/0/folders/15Wy1j8GB-l0ykpo4QGu3kUkraesLQRpc

**Enquiries:**
If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines)

Nordon Clementine (Orcid ID: 0000-0001-9441-9723)

## Title page

## Title

A first step towards best practice recommendations for the design and statistical analyses of pragmatic clinical trials: a modified Delphi approach

## Running title

Statistical analyses of pragmatic trials

## Authors

Victoria Palin[1], Tjeerd P. Van Staa[1], Stephanie Steels[2], Andrea B. Troxel[3], Rolf H.H. Groenwold[4], Tom M. MacDonald[5], David Torgerson[6], Douglas Faries[7], Pierre Mancini[8], Mario Ouwens[9], Lucy J. Frith[10], Kate Tsirtsonis[11], Graham MacLennan[12], and Clementine Nordon[13] on behalf of the GetReal Initiative

## Affiliations

[1] victoria.palin@manchester.ac.uk and tjeerd.vanstaa@manchester.ac.uk, Division of Informatics, Imaging & Data Sciences, Manchester Environmental Research Institute, University of Manchester, United Kingdom

[2] S.Steels@mmu.ac.uk, Department of Social Care and Social Work, Manchester Metropolitan University, Manchester M15 6GX, United Kingdom

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/bcp.15441

This article is protected by copyright. All rights reserved.

[3] Andrea.Troxel@nyulangone.org, Division of Biostatistics, Department of Population Health, NYU Grossman School of Medicine, NYU, USA

[4] R.H.H.Groenwold@lumc.nl, Department of Clinical Epidemiology, Leiden University Medical Centre, The Netherlands

[5] t.m.macdonald@dundee.ac.uk, MEMO Research, University of Dundee, Ninewells Hospital & Medical School, Dundee, DD1 9SY, United Kingdom

[6] david.torgerson@york.ac.uk, Department of Health Sciences, University of York, United Kingdom

[7] faries_douglas_e@lilly.com, Global Statistical Sciences, Eli Lilly & Co., Indianapolis IN, USA

[8] Pierre.Mancini@sanofi.com, Sanofi R&D, Chilly Mazarin, France

[9] Mario.Ouwens@astrazeneca.com, AstraZeneca, KC6, Gothenburg, Sweden

[10] lucy.j.frith@gsk.com, GlaxoSmithKline, Brentford, TW8 9GS, United Kingdom

[11] kbendall@amgen.com, Amgen Ltd, Uxbridge, United Kingdom

[12] g.maclennan@abdn.ac.uk, The Centre for Healthcare Randomised Trials, University of Aberdeen, United Kingdom

[13] nordon.clementine@gmail.com, formally LASER Research, Paris, France; currently AstraZeneca, Cambridge, United Kingdom

## Corresponding author

Dr Clementine Nordon, MD, PhD

nordon.clementine@gmail.com

## Key words

Pragmatic clinical trials; cluster-randomized trials; effectiveness; routine clinical practice; bias

Accepted Article

This article is protected by copyright. All rights reserved.

## Word count

Abstract: 243

Main text: 3,064

2 Tables

1 Figure

## Bullet point summary

### What is already known about this subject

- Randomised controlled trials (RCTs) usually adopt a highly controlled and experimental setting to confirm a causal relationship between intervention and outcome;
- Pragmatic clinical trials (PCTs) are RCTs implemented through routine clinical practice to increase generalizability of results and inform decision-makers on the effect of medical interventions in routine clinical practice; consequently, the design features of PCTs are less controlled
- However, modifying the design features of trials to improve pragmatism introduces statistical issues;

### What this study adds

- Best practice recommendations were generated by an international panel of experts providing a general framework for trialists to identify risks of bias in PCTs;
- These potential risks need to be gauged in the light of the trial estimand and can be minimized primarily by design, and/or through data analyses;

Accepted Article

This article is protected by copyright. All rights reserved.

## Abstract

**Aim** Pragmatic clinical trials (PCTs) are randomised trials implemented through routine clinical practice, where design parameters of traditional randomised controlled trials are modified to increase generalizability. However, this may introduce statistical challenges. We aimed to identify these challenges and discuss possible solutions leading to best practice recommendations for the design and analysis of PCTs.

**Methods** A modified Delphi method was used to reach consensus among a panel of 11 experts in clinical trials and statistics. Statistical issues were identified in a focused literature review and aggregated with insights and possible solutions from expert collected through a series of survey iterations. Issues were ranked according to their importance.

**Results** 27 articles were included and combined with experts' insight to generate a list of issues categorized into: participants; recruiting sites; randomisation, blinding and intervention; outcome (selection and measurement); and data analysis. Consensus was reached about the most important issues: risk of participants' attrition; heterogeneity of "usual care" across sites; absence of blinding; use of a subjective endpoint; and data analysis aligned with the trial estimand. Potential issues should be anticipated and preferably be addressed in the trial protocol. The experts provided solutions regarding data collection and data analysis, which were considered of equal importance.

**Discussion** A set of important statistical issues in PCTs was identified and approaches were suggested to anticipate and/or minimize these through data analysis. Any impact of choosing a pragmatic design feature should be gauged in the light of the trial estimand.

Accepted Article

This article is protected by copyright. All rights reserved.

## Main text

### Introduction

Randomised controlled trials (RCTs) are widely used for the evaluation of medical interventions such as medicines. As has been recognised several decades ago, the research questions of RCT can vary from explanatory to more pragmatic ones [1]. In explanatory RCTs, the research question concerns the testing of a biological mechanism and evaluation of causal effects of e.g. medicines. This involves careful selection of study participants most likely to respond to treatment and careful monitoring of study participants in order to ensure that the intervention is applied as intended and study participants take the medicines. Explanatory RCTs are typically used to obtain marketing approval for the medicine. In contrast, pragmatic RCTs test the effects of an intervention in usual clinical practices, whether the decision to use the intervention improves health outcomes in the usual healthcare delivery process [1].

These RCTs typically are closer to the routine clinical practice in terms of e.g., participants (broader edibility criteria), recruiting sites (representative of routine healthcare providers), patient clinical assessment and follow-up, choice of comparator, or of outcome [2, 3]. There is a continuum between the explanatory attitude of an RCT and the pragmatic attitude of a PCT and the extent to which a trial is pragmatic can be assessed using the PRECIS [4] and PRECIS-2 tool [5]. This tool scores a trial design on nine different domains in order to assess whether the trial design matches the intended use of the trial results however it does not provide a detailed overview of the statistical challenges associated with PCTs.

Modifying the design features of trials to improve pragmatism introduces various challenges related to their implementation [6] or to compliance with good clinical practice [7]. Moreover, these modifications introduce statistical issues that remain only partially explored

This article is protected by copyright. All rights reserved.

Accepted Article

or solved [8-11]. No guidelines have been developed in this research area that would involve a wide range of experts. To bridge this gap, we set out to identify and prioritize the statistical issues incurred by the design and conduct of PCTs, and where possible suggest potential statistical solutions to address them, and develop best practice recommendations using a panel of international experts in clinical trials and statistics. This study was conducted in the realm of the GetReal Consortium [12] which brought together stakeholder groups involved in the generation, analyses, and use of real-world data for drug development (i.e., pharmaceutical industry, academia, health technology and regulatory bodies) in order to improve and promote the generation of real-world data earlier in the drug development process. It included an almost even balance of participants from the public sector and the private sector.

## Methods

The Delphi (and Modified Delphi) technique is a qualitative research method used to build consensus amongst a group of experts around complex issues [13, 14]. The technique uses a series of rounds or iterations where experts independently provide information on a given topic.

Our study used aspects of classical and modified Delphi techniques [15] to identify the potential statistical issues present in pharmacological PCTs and, where possible solutions based on experts' experience in conducting PCTs, reaching consensus among experts so as to generate best practice recommendations. The material generated was processed, thematically analysed, and summarised by three researchers independent of the expert panel (VP, SS, and CN).

This article is protected by copyright. All rights reserved.

## Phase 1 - Expert Panel Recruitment

Ethical approval was sought from The University of Manchester ethics committee which considered a formal ethical review unnecessary because the study would ask professionals questions strictly within their professional competence.

We identified experts using the following criteria: (i) epidemiologists or statisticians, (ii) working in academia or in the pharmaceutical industry, (iii) with extensive experience in designing and/or analysing PCT data as per their publication records. A minimum of eight experts was deemed necessary to providing sufficient insight and expertise [16]. We also sought diversity in terms of country of origin, and as far as possible, gender. Thirty eight experts (24 academics and 14 industry professionals) from academia working in trial design, e.g., directors in pragmatic trials units, and industry trial experts participating to the Getreal Consortium were identified and invited to take part via email of whom 11 agreed to participate. For more information regarding the experts' area of expertise, see supplementary Table S1.

## Phase 2 – Focused Literature Review

The first source of information was derived from a focused literature review to identify the statistical impact of modifying traditional RCT features to make them more pragmatic (hereafter called statistical issues), and potential solutions to overcome these issues. The search algorithm (supplementary Table S2) was run in PubMed and Embase. Twenty-five articles were identified and screened independently by the study researchers on Title and Abstract, resulting in 18 articles that were read in full and 13 included. The reference lists of included articles were reviewed for further relevant articles, providing 14 additional articles and a final set of 27 articles (identified as of 2019; supplementary Table S3). The set of statistical issues identified were categorised into five design parameter categories: (1)

This article is protected by copyright. All rights reserved.

Participants; (2) Recruiting sites; (3) Randomisation, blinding and intervention; (4) Outcome selection and measurement; and (5) Data analyses.

### Phase 3 – Iterations of Expert Review and discussions

The iteration process of data generation and consensus building among experts is detailed in Figure 1. All surveys were circulated via email and replies were sent to the researcher team independent of other experts. The first few iterations followed a classical Delphi approach, whereby open qualitative information was gathered from multiple sources to gain a comprehensive list of known statistical issues in PCTs, and any possible known analytical solutions. The first survey asked experts to independently list the main issues and solutions, if available, within each of the above-described design categories. The results from the focused literature review and from the first survey were presented to the experts during a day-long in person meeting in order to gather their feedback and further insight. These sources of qualitative information were pooled, analysed thematically and synthesised, thus providing material for a second survey.

A modified Delphi process was then adopted for alternative rounds of independent surveys and virtual group discussions (2 hours each). This approach was adopted to consolidate and refine the data generated (e.g. clarification on reasoning or to refine discrepancies in terminology). The second survey asked experts to independently rank the perceived levels of importance of the impact that each issue may have on the validity of PCT results, as well as to provide and/or review the current recommended solutions. The ranking from experts was collated to generate a global ranking for each issue within each design category, whilst also identifying the level of disagreement between experts to be discussed in later iterations. The first virtual meeting aimed to clarify any reason for disagreement and reduce the number of issues down to five for each design category (thus a total of 25 issues). Due to the volume of issues raised, it was decided with the experts to focus on the top five for each design

This article is protected by copyright. All rights reserved.

parameter, for each category, to focus on the most important potential issues that trialists and clinicians should bear in mind when designing PCTs, therefore statistical issues that did not receive a ranking were deemed less important and disregarded for future iterations.

The third survey presented the ten most important issues by design category and where experts were asked to independently review and comment on the newly synthesised list of statistical issues and their level of potential impact on the validity of PCT results. The results from this survey were then analysed and discussed with experts during a final teleconferences. The second discussion focused on reviewing the statistical solutions. The aim was to explore all possible solutions available, regardless of agreement among experts.

Following these iterations, a set of Best Practice Recommendations for the statistical analysis of PCTs were sent to the experts for their final review and validation.

## Results

Regarding surveys, seven experts provided responses to survey one, nine experts to survey two, and nine experts to survey three. All eleven experts attended the in person meeting and teleconferences and provided feedback on the final report.

The results of the focused literature review, outputs from the first survey, and additional insight gained during the in person meeting resulted in a list of 85 statistical issues. After the third survey, saturation of issues was reached, i.e., no additional issues and solutions were reported in future discussion with experts.

### Most important statistical issues

The most important statistical issues within each design category are listed in Table 1. The issues related to “Participants” were: willingness to stay in the trial and attrition; absence of adherence-enhancement strategy; broad eligibility criteria and heterogeneity of treatment

This article is protected by copyright. All rights reserved.

Accepted Article

effects; absence of a standard and strict definition of disease, and motivation to take part in the trial. Regarding "Recruiting sites", the key issues were due to the heterogeneity of care provided; heterogeneity in the availability of treatment strategies; heterogeneity of data measurement and collection; heterogeneity in the ability to take part in the trial; and fragmentation of care. Design features related to "Randomisation, blinding and intervention" and leading to potential statistical issues included: the absence of blinding; treatment switching between treatment arms; physicians' preference and subversion; and selection bias in cohort-multiple RCTs (a type of pragmatic design where participants are sourced and randomised from a large cohort of patients) which is differential between the intervention and the control arm. In terms of selection and measurement of outcome in PCTs, the choice of a subjective endpoint, the difficulties in collecting outcome data, and the timing of outcome measurement were deemed important by experts. Finally, regarding data analyses of PCTs, the challenges included: analyses being aligned with the estimand targeted for the treatment effect, the handling of missing outcome data and the lack of statistical power. Most importantly, the experts emphasized that the set up design and planned analysis of trial data were essential and needed scrutinising in the design phase, ensuring all aspects of the trial aligned with the estimand.

Over the process of reaching an agreement, the perceived importance of issues was commonly agreed for the vast majority. For a few issues though, the experts expressed differing perspectives even after clarification of any misunderstanding. For example, there was some disagreement on whether including participants regardless of their level of adherence to treatment (or the absence of adherence-enhancement strategy) was an issue. One perspective held was that different levels of adherence to treatment increase the risk of misclassification of exposure. The second perspective was that adherence to treatment is an intrinsic aspect of treatments' effectiveness in routine clinical practice, estimation of which is

This article is protected by copyright. All rights reserved.

the purpose of PCTs. Through the discussions related to further understanding discrepancies in perceiving the importance of specific issues, including the one above, the experts agreed that the impact of a specific statistical issue actually depends on the trial estimand, i.e., the target of estimation. For instance, the absence of adherence-enhancement strategy is an issue if the aim of the PCT is to explore treatment safety; in this case adherence to treatment should be ascertained to be able to attribute any adverse event to the actual exposure to treatment. If the aim of the PCT is to measure treatments' effectiveness of which adherence is one aspect, then any lack of adherence should not be perceived as an issue but rather an aspect of treatment to be measured, e.g., through self-reported questionnaires or by the medication possession ratio.

### Solutions to mitigate statistical issues

A large array of solutions was provided by experts without any prioritisation between them. Early during the course of discussions, the experts highlighted the importance of identifying possible issues early during the trial design process and mitigating them at a protocol-level. Solutions based on analytical approaches (e.g., imputation techniques of missing data) were deemed as important as strategies that can be implemented before and during the trial conduct (e.g., assessing the completeness and quality of data collection in sites prior to trial start-up). The solutions were therefore categorized into "set up solutions" (how to design the PCT so as to minimise any risk of bias) and "analytical solutions" (how to account for potential bias in the statistical analyses). These solutions are listed in Table 2. Regarding set up solutions, the experts stressed the importance of planning the collection of adequate data allowing for the conduct of specific analyses to identify bias, e.g., measuring the reasons for patients' attrition and switching, measuring adherence level, collecting characteristics of sites, measuring patients' expectation of treatment. Regarding data analyses, the experts emphasised the importance of carefully describing trial data before using more complex analyses.

This article is protected by copyright. All rights reserved.

## Discussion

In this study, we thoroughly identified the potential statistical issues and risk of bias incurred by the use of pragmatic design features in a clinical trial. Various sources of information were combined (focused literature review and expert insight) and synthetized. A consensus was built among 11 experts on the most critical issues and listed along with possible set up or analytical solutions to mitigate them.

## Key results

The importance of clearly defining the trial estimand appeared to be the key starting point for conducting a good-quality trial. The necessity for experts to discuss extensively the importance of some issues led to the conclusion that "this all depends on the trial estimand". In the light of the trial estimand, an issue could be more or less of a concern. Although the design of a clinical trial and the question that it intends to address seem to be straightforward (what is the treatment *effect*? "Does the intervention *work* under ideal circumstance?" [17]), the specific measure behind what is called "treatment effect" is subtle and needs to be clarified to avoid misunderstanding and misinterpretation of results. The International Council for Harmonisation (ICH) of clinical trials has recently developed a framework using the construction of the estimand to facilitate precision in describing a treatment effect of interest [18].

Although the solutions were initially envisaged on the analytical angle, the experts emphasised the need for potential issues to be anticipated at a protocol level. Putting in place strategies to minimize bias before and during the conduct of a trial reduces the need for complex analyses to account for it. These set up solutions included, for instance, the collection of data necessary to identify the source of a possible bias, or suggestions to

This article is protected by copyright. All rights reserved.

improve the quality of outcome measurement or data collection in recruiting sites that are less used to participate in trials.

Altogether, the best practice recommendations provided by experts follow a step-by-step approach: (1) defining precisely the trial estimand; (2) anticipating possible statistical issues during the protocol development; (3) planning for set up solutions to be implemented before or during the trial conduct; (4) carefully describing any source of bias using trial data; and if necessary (5) performing more complex statistical analyses to adjust for these biases.

In the literature, a large proportion of the challenges and sources of bias in PCTs that were identified and discussed are related to cluster randomized trials (CRTs) and the risk of insufficient statistical power. Regarding CRTs, post-randomisation recruitment and subversion were identified as causing potential differential recruitment patterns and imbalance of participants' baseline characteristics [19-23]. In the present study, the experts also considered post-randomisation recruitment in CRTs as being critical. Previous studies reported other issues specific to CRTs: intra-cluster correlation [11], "unit-of-analysis error" overestimating treatment effect size [24], risk of imbalanced participants characteristics at baseline [25] and risk of unequal cluster size leading to reduced statistical power [10]. Although these issues were listed and discussed by experts, they were not considered as being the most impactful; moreover, these issues are specific to CRTs. Reasons for potential insufficient statistical power include the recruitment of a more heterogeneous population and the inability to detect a treatment-by-subgroup interaction [9, 26], the conduct of stepped-wedge cluster randomized trials [27] or of cohort-multiple RCTs [28]. In the present study, the risk of not reaching adequate statistical power was considered important, but for reasons that are numerous and go beyond the conduct of a CRT or a stepped-wedge CRT. Contrary to previous study reports, the experts listed and discussed issues and solutions irrespective of a specific trial design, thus generating a general framework applicable to any type of PCT.

This article is protected by copyright. All rights reserved.

## Strength and Limitations

A diverse panel of experts from several countries took part in the present study. Experts were academic researchers in the realm of clinical trial methodology and pharmaceutical industry experts bringing specific insight on the use of PCTs for regulatory purposes. The variety in experts' background provided different and complementary perspectives on trial methodologies. Participation of experts was good and constant throughout the study conduct. Moreover, saturation of themes (issues and solutions) was reached after the last survey. Altogether, we believe that the set of best practice recommendations generated reflects a thorough and complete picture of the statistical consequences and hurdles incurred by the conduct of PCTs.

One limitation of our study is that the design features are not fully aligned with the PRECIS-2 framework that consists in nine dimensions, and not five design categories. This may be confusing for readers familiar with this framework. However, there is an overlap between our categorization and the PRECIS-2 framework and the issues listed by experts all fall into one of the nine PRECIS-2 dimensions. For instance, three PRECIS-2 dimensions correspond to our design category "recruiting sites": "recruitment" (how are participants recruited in the trial?), "setting" (where is the trial being done?) and "organisation" (what expertise and resources are needed to deliver the intervention?). Some statistical issues identified as crucial for experts go beyond the PRECIS-2 framework, e.g., missing outcome data, because the latter aims at exploring the level of pragmatism of a PCT and not at anticipating specific statistical issues related to this pragmatism. Another limitation was that our study focused on identifying statistical issues in PCTs for pharmacological interventions; our results may be generalizable to other types of medical interventions (e.g., surgery, therapeutic education, psychotherapies) but this was not explored.

This article is protected by copyright. All rights reserved.

## Implication and conclusion

The best practice recommendations generated by this study provide a general framework for trialists to identify risks of bias and minimize them by design or through data analyses. This framework aims at bringing awareness on the key questions that need to be discussed and addressed when designing a trial. The outputs of the present study should be considered as a starting point of discussions and future research involving the broader scientific community.


## Acknowledgements

The authors did not preregister the research in an independent institutional registry.


## Conflict of interest statement

No author has any COI to declare with regards this purely methodological study. D. Faries is an employee at Eli Lilly, P. Mancini is an employee at Sanofi, M. Ouwens is an employee at AstraZeneca, L. Frith is an employee at GSK, K. Tsirtsonis is an employee at Amgen. C. Nordon was an employee at LASER Research at the time of the study conduct and is now an employee at AstraZeneca.


## Funding information

The study was funded by the IMI program, grant agreement number 807012. No expert from Academia was compensated for the time spent in the study, including meetings.



Accepted Article

This article is protected by copyright. All rights reserved.

## Data availability statement

No Human subject participated in the present study. The data generated and analysed during the current study are available in the Google Drive repository without any restriction: https://drive.google.com/drive/u/0/folders/15Wy1j8GB-l0ykpo4QGu3kUkraesLQRpc

## Author contribution statement

- Conception or design of the work: Tjeerd van Staa, Victoria Palin, Stephanie Steels and Clementine Nordon
- Data collection: Victoria Palin, Tjeerd P. Van Staa, Stephanie Steels, Andrea B. Troxel, Rolf H.H. Groenwold, Tom M. MacDonald, David Torgerson, Douglas Faries, Pierre Mancini, Mario Ouwens, Lucy J. Frith, Kate Tsirtsonis, Graham MacLennan, and Clementine Nordon; all authors contributed equally
- Data analysis and interpretation: Victoria Palin, Tjeerd P. Van Staa, Stephanie Steels, Andrea B. Troxel, Rolf H.H. Groenwold, Tom M. MacDonald, David Torgerson, Douglas Faries, Pierre Mancini, Mario Ouwens, Lucy J. Frith, Kate Tsirtsonis, Graham MacLennan, and Clementine Nordon; all authors contributed equally
- Drafting the article: Victoria Palin and Clementine Nordon, with input from all authors
- Critical revision of the article: Victoria Palin, Tjeerd P. Van Staa, Stephanie Steels, Andrea B. Troxel, Rolf H.H. Groenwold, Tom M. MacDonald, David Torgerson, Douglas Faries, Pierre Mancini, Mario Ouwens, Lucy J. Frith, Kate Tsirtsonis, Graham MacLennan, and Clementine Nordon; all authors contributed equally
- Final approval of the version to be published: Victoria Palin, Tjeerd P. Van Staa, Stephanie Steels, Andrea B. Troxel, Rolf H.H. Groenwold, Tom M. MacDonald,

This article is protected by copyright. All rights reserved.

Accepted Article

David Torgerson, Douglas Faries, Pierre Mancini, Mario Ouwens, Lucy J. Frith, Kate Tsirtsonis, Graham MacLennan, and Clementine Nordon;

Clementine Nordon is guarantor.

## References


1. Schwartz, D. and J. Lellouch, *Explanatory and pragmatic attitudes in therapeutical trials.* J Chronic Dis, 1967. **20**(8): p. 637-48.
2. Zwarenstein, M., et al., *Improving the reporting of pragmatic trials: an extension of the CONSORT statement.* BMJ, 2008. **337**: p. a2390.
3. Ford, I. and J. Norrie, *Pragmatic Trials.* N Engl J Med, 2016. **375**(5): p. 454-63.
4. Thorpe, K.E., et al., *A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers.* J Clin Epidemiol, 2009. **62**(5): p. 464-75.
5. Loudon, K., et al., *The PRECIS-2 tool: designing trials that are fit for purpose.* BMJ, 2015. **350**: p. h2147.
6. Zuidgeest, M.G.P., et al., *Series: Pragmatic trials and real world evidence: Paper 1. Introduction.* J Clin Epidemiol, 2017. **88**: p. 7-13.
7. Mentz, R.J., et al., *Good Clinical Practice Guidance and Pragmatic Clinical Trials: Balancing the Best of Both Worlds.* Circulation, 2016. **133**(9): p. 872-80.
8. Gamerman, V., Cai, T., Elsäßer, A. , *Pragmatic randomized clinical trials: best practices and statistical guidance.* Health Serv Outcomes Res Method, 2019. **19**: p. 23-35.
9. Troxel, A.B., D.A. Asch, and K.G. Volpp, *Statistical issues in pragmatic trials of behavioral economic interventions.* Clin Trials, 2016. **13**(5): p. 478-83.
10. Cook, A.J., et al., *Statistical lessons learned for designing cluster randomized pragmatic clinical trials from the NIH Health Care Systems Collaboratory Biostatistics and Design Core.* Clin Trials, 2016. **13**(5): p. 504-12.
11. Califf, R.M., *Pragmatic clinical trials: Emerging challenges and new roles for statisticians.* Clin Trials, 2016. **13**(5): p. 471-7.
12. Innovative Medicines Initiative. *The GetReal Consortium*. 2013; Available from: http://www.imi-getreal.eu/.
13. The Rand Corporation. *Delphi Method*. Available from: https://www.rand.org/topics/delphi-method.html.
14. McKenna, H.P., *The Delphi technique: a worthwhile research approach for nursing?* J Adv Nurs, 1994. **19**(6): p. 1221-5.
15. Hasson, F. and S. Keeney, *Enhancing rigour in the Delphi technique research.* Technological Forecasting and Social Change, 2011. **78**(9): p. 1695-1704.
16. Hallowell, M.R. and J.A. Gambatese, *Qualitative Research: Application of the Delphi Method to CEM Research.* Journal of Construction Engineering and Management, 2010. **136**(1): p. 99–107.
17. Singal, A.G., P.D. Higgins, and A.K. Waljee, *A primer on effectiveness and efficacy trials.* Clin Transl Gastroenterol, 2014. **5**: p. e45.
18. International council for harmonisation. *Addendum on estimands and sensitivity analysis in clinical trials To the guideline on statistical principles for clinical trials*


This article is protected by copyright. All rights reserved.

2019 [cited 2021 October 5th ]; Available from: https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf.
19. Hahn, S., et al., *Methodological bias in cluster randomised trials.* BMC Med Res Methodol, 2005. **5**: p. 10.
20. Giraudeau, B. and P. Ravaud, *Preventing bias in cluster randomised trials.* PLoS Med, 2009. **6**(5): p. e1000065.
21. Pence, B.W., et al., *Balancing Contamination and Referral Bias in a Randomized Clinical Trial: An Application of Pseudo-Cluster Randomization.* Am J Epidemiol, 2015. **182**(12): p. 1039-46.
22. Barnett, L.A., et al., *Applying quantitative bias analysis to estimate the plausible effects of selection bias in a cluster randomised controlled trial: secondary analysis of the Primary care Osteoarthritis Screening Trial (POST).* Trials, 2017. **18**(1): p. 585.
23. Dickinson, L.M., et al., *Pragmatic Cluster Randomized Trials Using Covariate Constrained Randomization: A Method for Practice-based Research Networks (PBRNs).* J Am Board Fam Med, 2015. **28**(5): p. 663-72.
24. Eldridge, S.M., et al., *Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care.* Clin Trials, 2004. **1**(1): p. 80-90.
25. Bolzern, J., Mnyaman N., Bosanquet, K., Torgerson, D., *Comparing evidence of selection bias between cluster-randomised and individually randomised controlled trials: a systematic review and meta-analysis*, in *The Lancet*. 2018.
26. Dember, L.M., et al., *Pragmatic Trials in Maintenance Dialysis: Perspectives from the Kidney Health Initiative.* J Am Soc Nephrol, 2016. **27**(10): p. 2955-2963.
27. Eichner, F.A., et al., *Systematic review showed that stepped-wedge cluster randomized trials often did not reach their planned sample size.* J Clin Epidemiol, 2019. **107**: p. 89-100.
28. Candlish, J., et al., *Evaluation of biases present in the cohort multiple randomised controlled trial design: a simulation study.* BMC Med Res Methodol, 2017. **17**(1): p. 17.
29. Chhatre, S., et al., *Patient-centered recruitment and retention for a randomized controlled study.* Trials, 2018. **19**(1): p. 205.
30. Nordon, C., et al., *Trial exclusion criteria and their impact on the estimation of antipsychotic drugs effect: A case study using the SOHO database.* Schizophr Res, 2018. **193**: p. 146-153.
31. Pocock, S.J. and G.W. Stone, *The Primary Outcome Fails - What Next?* N Engl J Med, 2016. **375**(9): p. 861-70.
32. Chassang, S., et al., *Accounting for Behavior in Treatment Effects: New Applications for Blind Trials.* PLoS One, 2015. **10**(6): p. e0127227.
33. Kravitz, R.L., N. Duan, and J. Braslow, *Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages.* Milbank Q, 2004. **82**(4): p. 661-87.
34. Richman, J.S., S. Andreae, and M.M. Safford, *Challenges of Prolonged Follow-up and Temporal Imbalance in Pragmatic Trials: Analysis of the ENCOURAGE Trial.* Ann Fam Med, 2015. **13 Suppl 1**: p. S66-72.
35. MacDonald, T.M., et al., *Randomized trial of switching from prescribed non-selective non-steroidal anti-inflammatory drugs to prescribed celecoxib: the Standard care vs. Celecoxib Outcome Trial (SCOT).* Eur Heart J, 2017. **38**(23): p. 1843-1850.
36. Christian, J.B., et al., *Masking in Pragmatic Trials: Who, What, and When to Blind.* Ther Innov Regul Sci, 2020. **54**(2): p. 431-436.
37. Stroup, T.S., et al., *The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project: schizophrenia trial design and protocol development.* Schizophr Bull, 2003. **29**(1): p. 15-31.

This article is protected by copyright. All rights reserved.

38. Pate, A., et al., *Cohort Multiple Randomised Controlled Trials (cmRCT) design: efficient but biased? A simulation study to evaluate the feasibility of the Cluster cmRCT design.* BMC Med Res Methodol, 2016. **16**(1): p. 109.
39. Kingsbury, S.R., et al., *Pain reduction with oral methotrexate in knee osteoarthritis, a pragmatic phase iii trial of treatment effectiveness (PROMOTE): study protocol for a randomized controlled trial.* Trials, 2015. **16**: p. 77.
40. Almario, C.V., et al., *Impact of National Institutes of Health Gastrointestinal PROMIS Measures in Clinical Practice: Results of a Multicenter Controlled Trial.* Am J Gastroenterol, 2016. **111**(11): p. 1546-1556.
41. Hartman, L., et al., *Harm, benefit and costs associated with low-dose glucocorticoids added to the treatment strategies for rheumatoid arthritis in elderly patients (GLORIA trial): study protocol for a randomised controlled trial.* Trials, 2018. **19**(1): p. 67.
42. Holmes, L., et al., *Innovating public engagement and patient involvement through strategic collaboration and practice.* Res Involv Engagem, 2019. **5**: p. 30.
43. Tang, K.L., H. Quan, and D.M. Rabi, *Measuring medication adherence in patients with incident hypertension: a retrospective cohort study.* BMC Health Serv Res, 2017. **17**(1): p. 135.
44. Thompson, K., J. Kulkarni, and A.A. Sergejew, *Reliability and validity of a new Medication Adherence Rating Scale (MARS) for the psychoses.* Schizophr Res, 2000. **42**(3): p. 241-7.
45. Sanders, E., P. Gustafson, and M.E. Karim, *Incorporating partial adherence into the principal stratification analysis framework.* Stat Med, 2021. **40**(15): p. 3625-3644.
46. Dunn, G., M. Maracy, and B. Tomenson, *Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods.* Stat Methods Med Res, 2005. **14**(4): p. 369-95.
47. Robins, J.M., *Correction for non-compliance in equivalence trials.* Stat Med, 1998. **17**(3): p. 269-302; discussion 387-9.
48. Schmidt, A.F., et al., *Exploring interaction effects in small samples increases rates of false-positive and false-negative findings: results from a systematic review and simulation study.* J Clin Epidemiol, 2014. **67**(7): p. 821-9.
49. Pitt, B., et al., *Spironolactone for heart failure with preserved ejection fraction.* N Engl J Med, 2014. **370**(15): p. 1383-92.
50. Zhao, Y., et al., *Estimating Individualized Treatment Rules Using Outcome Weighted Learning.* J Am Stat Assoc, 2012. **107**(449): p. 1106-1118.
51. Liang, M., T. Ye, and H. Fu, *Estimating individualized optimal combination therapies through outcome weighted deep learning algorithms.* Stat Med, 2018. **37**(27): p. 3869-3886.
52. Singleton, P. and M. Wadsworth, *Consent for the use of personal medical data in research.* BMJ, 2006.
53. Happich, M., et al., *Reweighting Randomized Controlled Trial Evidence to Better Reflect Real Life - A Case Study of the Innovative Medicines Initiative.* Clin Pharmacol Ther, 2020. **108**(4): p. 817-825.
54. Simon, G.E., S.M. Shortreed, and L.L. DeBar, *Zelen design clinical trials: why, when, and how.* Trials, 2021. **22**(1): p. 541.
55. Leuchs, A.K., et al., *Disentangling estimands and the intention-to-treat principle.* Pharm Stat, 2017. **16**(1): p. 12-19.
56. Hernan, M.A. and J.M. Robins, *Per-Protocol Analyses of Pragmatic Trials.* N Engl J Med, 2017. **377**(14): p. 1391-1398.
57. Lash, T.L., et al., *Methods to apply probabilistic bias analysis to summary estimates of association.* Pharmacoepidemiol Drug Saf, 2010. **19**(6): p. 638-44.

This article is protected by copyright. All rights reserved.

58. Ring, A. and M.J. Wolfsegger, *The potential of the estimands framework for clinical pharmacology trials: Some discussion points.* Br J Clin Pharmacol, 2020. **86**(7): p. 1240-1247.
59. Lipkovich, I., B. Ratitch, and C.H. Mallinckrodt, *Causal Inference and Estimands in Clinical Trials.* Statistics in Biopharmaceutical Research, 2020. **12**(1): p. 54-67.
60. Cai, X., et al., *Estimands and missing data in clinical trials of chronic pain treatments: advances in design and analysis.* Pain, 2020. **161**(10): p. 2308-2320.
61. Psioda, M.A., et al., *Methodological Challenges and Statistical Approaches in the COMprehensive Post-Acute Stroke Services Study.* Med Care, 2021. **59**(Suppl 4): p. S355-S363.
62. Eldridge, S., et al. *Revised Cochrane risk of bias tool for randomized trials (RoB 2.0): Additional considerations for cluster randomized trials*. 2016 [cited 2022 April 28th]; Available from: https://www.unisa.edu.au/contentassets/72bf75606a2b4abcaf7f17404af374ad/rob2-0_cluster_parallel_guidance.pdf.
63. Hemming, K., et al., *The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting.* BMJ, 2015. **350**: p. h391.
64. Reeves, D., et al., *The cohort multiple randomized controlled trial design was found to be highly susceptible to low statistical power and internal validity biases.* J Clin Epidemiol, 2018. **95**: p. 111-119.
65. Turkoz, I., M. Sobel, and L. Alphs, *Application of Bayesian analyses to doubly randomized delayed start, matched control designs to demonstrate disease modification.* Pharm Stat, 2019. **18**(1): p. 22-38.
66. Tyson, J.E., et al., *Stopping guidelines for an effectiveness trial: what should the protocol specify?* Trials, 2016. **17**(1): p. 240.
67. Yang, R., et al., *Selection bias and subject refusal in a cluster-randomized controlled trial.* BMC Med Res Methodol, 2017. **17**(1): p. 94.
68. Yuan, A., et al., *Subgroup analysis with semiparametric models toward precision medicine.* Stat Med, 2018. **37**(11): p. 1830-1845.
69. Zhan, Z., G.H. de Bock, and E.R. van den Heuvel, *Statistical methods for unidirectional switch designs: Past, present, and future.* Stat Methods Med Res, 2018. **27**(9): p. 2872-2882.

Accepted Article

This article is protected by copyright. All rights reserved.

**Table 1:** Important statistical issues in pragmatic clinical trials as per experts' opinion and categorized by study design parameter

| Modification of design feature | Statistical consequence |
|---|---|
| **1. Participants recruitment and follow-up** | |
| **1.1. Willingness to stay in the trial and attrition** | In PCTs, no specific method is used to improve participants' compliance to the trial protocol, in particular regarding the duration of the trial [29]. The willingness of participants to stay in the trial throughout its duration may affect their participation thus leading to attrition and missing outcome data. In addition, participants who are lost to follow-up are likely to have distinct characteristics, including worse functional outcomes, medical complications or death. |
| **1.2. Absence of adherence-enhancement strategy** | Contrary to numerous RCTs [30] the expected level of adherence to treatment is not an exclusion criterion in PCTs. Participants' level of adherence will be reflecting that of routine clinical practice.<br>The experts expressed two perspectives regarding this possible issue. On the one hand, lack of adherence may lead to erroneously considering participants as exposed thus leading to misclassification and dilution of the treatment effect [31]. On the other hand, adherence may be considered as an intrinsic aspect of treatment potentially reflecting its effectiveness and tolerability. The extent to which adherence is an issue will depend on the trial estimand. |
| **1.3. Broad eligibility criteria and heterogeneity of treatment effects** | As participants of heterogeneous characteristics are included in PCTs, it is possible that some phenotypes modifying the effect of treatment – treatment effect modifiers – are at play, e.g., older age, behavioural risk factors [32].<br>Heterogeneity of treatment effects reflects patients' diversity in disease risk, responsiveness to treatment, vulnerability to adverse effects, and utility for different outcomes [33]. However, in case of treatment effect modification, estimating only the average treatment effect is disputable because this assumes a similar treatment effect across heterogeneous patient characteristics. One consequence is a decrease of statistical power. Similarly, if treatment heterogeneity leads to more harm than benefit in some subgroups, the net average treatment effect may be close to zero and the trial may be negative, despite the potential benefit in some subgroups [33]. |
| **1.4. Absence of a standard and strict definition of disease** | In PCTs, the diagnosis of the disease of interest may be left at the discretion of recruiting physicians. In the absence of standard and consistent definition of the disease across sites, participants with a wide range of disease severity or even with a |

This article is protected by copyright. All rights reserved.

| | |
|---|---|
| | wrong diagnosis may be included in the trial. In this case, the treatment effect is likely to be underestimated, with a lack of statistical power to show the average treatment effect or to detect a treatment-by-subgroup interaction [3, 26, 27]. Heterogeneity of treatment effect is different from heterogeneity of patients however, the two are related. |
| **1.5. Motivation to take part in the trial and generalizability of results** | Recruiting participants who are representative of routine clinical care is one important aspect of PCTs. Some subjects will agree to participate in the trial while some will not, thus leading to self-selection bias.<br>Although this issue is not specific to PCTs, some experts stressed that self-selection of certain categories of subject is a direct violation of the very principle of PCTs. |
| **2. Recruiting sites** | |
| **2.1. Heterogeneity of care provided by participating sites** | Sites involved in the trial will have heterogeneous standards of "usual care" (e.g., clinical expertise, quality of medical care and process, providing the treatment in the right way, following best practice, having the adequate equipment).<br>The heterogeneity of sites participating in the trial may affect the extent to which the treatment will "work", because the treatment effect is "confounded" by other aspects of care. |
| **2.2. Heterogeneity in the availability of treatment strategies** | In addition, all eligible sites may not deliver the intervention of interest or the comparative treatment planned in the design of the trial. These sites may thus be unable to take part in the trial. |
| **2.3. Heterogeneity of data measurement and collection** | In PCTs, data may be collected through electronic medical records (eHR), namely to minimize the burden for physicians. Different recruiting sites may collect data in a heterogeneous manner. The harmonisation of data collection, that is necessary for statistical analyses, may be jeopardized by: a) various medical data collection processes or data capture tools (software, format of data, etc.); b) different coding systems (e.g., ICD-10, Read codes); c) different terminologies (different medical terms can be used). Second, data accessibility and granularity (completeness) may differ across sites thus leading to missing data. Collecting trial data through eHRs compared to case report forms (as in traditional RCTs) is prone to measurement errors or missing data because data collected in eHRs are not purposefully collected for trials. In addition, data is not collected by trained study investigators but rather as part of routine clinical practice. |
| **2.4. Heterogeneity in the ability to take part in the trial** | Eligible sites are also heterogeneous in terms of experience in participating to a trial; the less experienced sites may be less willing to participate, thus leading to a selection bias towards more experienced and bigger sites.<br>The quality of data collection may be lower in less experienced |

This article is protected by copyright. All rights reserved.

| | |
|---|---|
| | sites too. In smaller recruiting sites, the number of eligible subjects and, in turn, the possible number of participants in the PCT may affect the speed of recruitment (leading to unplanned temporal imbalance in recruitment and follow-up) or small sample size. It was noted that stepped-wedge trials are prone to secular trend [34]. |
| **2.5. Fragmentation of care** | Because patients may be treated by different healthcare professionals who may not share information, important information necessary for the trial (eligibility criteria or outcome) may be missing for the investigator. If participants are followed up in multiple sites, there is a risk of missing information. |
| **3. Randomisation, blinding and intervention** | |
| **3.1. Absence of blinding and patient expectations** | In PCTs, the intervention may be provided in a non-blinded fashion either because the intervention cannot be blinded (e.g., a complex intervention, a surgical technique) or because this was deemed more pragmatic. Patients' expectations or preferences may be superior for one intervention and the "new" intervention may be perceived as more appealing or vice versa [35].<br>The absence of blinding may increase drop-out rate, increase treatment discontinuation or switching rates [3, 8] and affect outcome measurement in a differential manner between study arms [10, 36]. |
| **3.2. Cross-over between treatment arms, treatment switching** | Therapeutic switch is possible in PCTs and may be influenced by the perceived inefficacy by patients or physicians or tolerability of the treatment [37], in particular in the absence of blinding. This has several consequences:<br>• The participation of patients in the original treatment arm is shorter than planned<br>• Switchers are not comparable to non-switchers; assessing treatment effect only in non-switchers leads to biased results.<br>The estimand needs to be correctly defined, e.g., is the aim of the trial to measure the comparative effect of being assigned to / initiating a treatment, or of actually taking the treatment? In the latter case, treatment switching is problematic. |
| **3.3. Physicians preference and subversion** | In a cluster-randomised trial, randomisation is done at the site level and not the patient level. Physicians know to which therapeutic arm their clinical site was randomized to. If the recruitment of patients takes place after the randomisation of sites (post-randomisation recruitment), subversion may influence the decision of a physician to include a particular patient. Physicians' preference can lead to selection bias and confounding (unbalanced characteristics between treatment arms at baseline) [19, 21, 22, 34]. |

This article is protected by copyright. All rights reserved.

| | |
|---|---|
| **3.4. Selection bias in cohort-multiple RCTs being differential between the intervention and the control arm** | In a cohort-multiple RCT, there is a two-step consent: first for all patients when included in the cohort (consent for data collection) and second, for switching to the new treatment (only for patients randomised to this therapeutic arm).<br>This design is prone to differential selection bias: refusal to treatment is only present in the intervention arm, and this may lead to bias and reduce statistical power [28, 38]. |
| **4. Outcome selection and measurement** | |
| **4.1. Choice of a subjective endpoint** | In a PCT the primary endpoint may involve subjective measures possibly reported by patients – patients-reported outcomes (e.g., pain, health-related quality of life, satisfaction with care) because these could be more relevant for patients. If this measure is reflecting a concept, or a latent trait like in questionnaires, the "truth" may differ from the measure obtained hence measurement error.<br>Subjective measures are more susceptible to individual interpretation. For example, in a trial on arthritis therapy [39], the investigator could assess a combination of objective (e.g., biological results, CT scan) and subjective measures (e.g., pain). In routine clinical care physicians adapt treatments based on patients' perception and not only using hard endpoints. |
| **4.2. Difficulties in collecting outcome data, missing outcome data** | In PCTs the risk of missing outcome data is important [40, 41] due to a) sites and physicians: heterogeneous mode of data collection (item 2.3) and fragmentation of care (item 2.5) and b) patients: willingness to stay in the trial and drop out (item 1.1), timing of outcome measurement (item 4.3).<br>If the error in outcome measurement is substantially different between therapeutic arms, this leads to differential misclassification bias. |
| **4.3. Timing of outcome measurement** | Contrary to traditional RCTs in which the collection of medical information is scheduled at pre-specified dates, in PCTs, the date for outcome measurement may be unspecified. The outcome of interest is recorded e.g., only when/if the patient comes for a medical visit and attends the healthcare setting<br>Depending on disease severity (which may be affected by the intervention), patients have more frequent (or less) visits, some of which may be unscheduled. This can lead to differential outcome measurement and bias. |
| **5. Data analysis** | |
| **5.1. Analysis aligned with the estimand** | PCTs may address many different questions and thus potentially target different estimands. The importance of statistical analysis being aligned with the estimand is particularly salient in PCTs.<br>The impact of loosening traditional design detailed above may be more or less important depending on the estimand, e.g., in |

This article is protected by copyright. All rights reserved.

| | |
|---|---|
| | relation to:<br>• Willingness to stay in the trial and risk of attrition;<br>• Adherence;<br>• Heterogeneity of treatment effect;<br>• Generalisability of results;<br>• Treatment switching;<br>• Timing of outcome measurement |
| **5.3. Handling of missing outcome data** | The difficulties in collecting outcome data and the risk of missing outcome data are more important in PCTs than in RCTs due to the reasons provided above (item 4.2). Handling missing outcome data is particularly challenging. |
| **5.3. Lack of statistical power** | PCTs can fail to reach the required and anticipated statistical power [27, 28, 38] due to:<br>• Fewer patients included than planned, e.g., reduced number of sites participating in CRTs;<br>• Patient behaviour (e.g., adherence)<br>• Increased variance of treatment effect due to heterogeneous population<br>• Increased heterogeneity of patients (broader patient eligibility of real world studies);<br>• The varying number of treatments regimes available to patients in clinical practice;<br>• Increased variation in the outcome / endpoint;<br>• Sporadic timing of assessment/measurement recordings |

Accepted Article

This article is protected by copyright. All rights reserved.

**Table 2:** Set up and analytical solutions recommended, to prevent or address statistical issues identified in pragmatic clinical trials, categorized by study design parameter

| Source of issue | Set up | Analytical |
|---|---|---|
| **1. Participants recruitment and follow-up** | | |
| **1.1. Willingness to stay in the trial and attrition** | Patients engagement<br>• Minimize the burden to patients of participating in the trial;<br>• Consider remote consultations<br>• Involve the public and patient groups through Patient and Public Involvement and Engagement [42] early in the design stage; provide regular feedback on the trial (e.g., online information on recruitment);<br>• Select a motivating and relevant intervention/outcome to engage interest of patients;<br><br>Minimize the risk of missing outcome data<br>• Provide training to physicians (at an individual/site level) so that patients participating in the trial are followed-up through the entire trial duration, i.e., until outcome measurement irrespective of what happens after randomisation (e.g., treatment switching or discontinuation).<br>• Increase the ease of data collection, e.g., use user-friendly tools for outcome assessments, when appropriate (e.g., smart phone, connected watch, glucometer);<br>• Capture some endpoints in other data sources (e.g., death registries) or by linking data to follow-up on patients that have dropped-out (ethics dependent);<br>• Pay specific attention to trial duration in the absence of participant retention strategy; | First, this may be relevant to understand reasons for attrition:<br>• Consider qualitative evaluation to gain an understanding of why patients dropped-out;<br>• The reasons for attrition can be described<br><br>Then, analytical solutions to manage missing outcome data depend on the nature of missingness and trial estimand.<br>• See item 5.2 on missing outcome data |

This article is protected by copyright. All rights reserved.

| | | |
|---|---|---|
| | • Plan to monitor patients participation when high drop-out rates are expected<br><br>Specific data collection<br>• Collect accurate reasons for attrition and consider sensitivity analyses to address potential issues due to missing data.<br>• Plan to collect time-varying data/measures and reasons for attrition preferably close to the censoring time; | |
| **1.2. Absence of adherence-enhancement strategy** | Is adherence important?<br>• Provide a clear definition of the trial estimand and specify whether lack of adherence should be accounted for;<br><br>What is the level of adherence?<br>• Measure adherence carefully. There are many methods to measure adherence including: collecting information on drug dispensing, e.g., possession ratio [43]; asking patients to return drug packets; measuring blood drug concentration when appropriate;<br>• Provide a definition of measurements indicative of a "good" adherence and how to measure it, using a threshold or a continuous score, e.g., using the Medication Adherence Report Scale (MARS) score [44].<br><br>In any case, patients that stop treatment need to continue participating in the trial | Adjust for adherence with attention to the estimand (safety or effectiveness).<br><br>If lack of adherence is an issue, the analytical methods to adjust for adherence include:<br>• Principal stratification approaches, i.e., identify underlying strata and then compute causal effects within strata – patient-level adherence [45];<br>• Estimation of the Complier-Average Causal Effect (CACE) – measuring the impact of an intervention in a subgroup of the population – e.g., using instrumental variables – another variable (instrument) [46];<br><br>Of note, this is important to understand whether lack of adherence is random, or related to the treatment [47]. |
| **1.3. Broad eligibility criteria and heterogeneity of treatment** | Is heterogeneity an issue?<br>• Provide a clear definition of the trial estimand and specify whether one interest of the trial is to identify and measure | The statistical approach depends on whether or not hypothesis on treatment effect modification was made<br>• When hypothesis were made on |

This article is protected by copyright. All rights reserved.

Accepted Article

| | | |
|---|---|---|
| **effects** | heterogeneity of treatment effects<br>• If so, sample size should be determined such that treatment effect heterogeneity can be detected. If the trial sample is to small (lack of statistical power) or the participants not representative of the actually treated patients, the trial may fail to show heterogeneity of treatment effect [48]. For instance, sicker patients or more adherent patients will be more likely to participate in the trial, while severity of disease or adherence may modify the effect of the drug in real world.<br><br>Measuring heterogeneity<br>• Explore the risk of effect modification, identify possible effect-modifiers and determine which variables should be collected to measure effect modification;<br>• Ideally, generate hypothesis on confounding and effect modification prior to data collection and data analyses.<br>• Establish biological or physiological rationale when considering effect modification: is it true effect modification? Is it spurious?<br><br>Measuring the average treatment effect<br>If it is relevant clinically to measure the average treatment effect, despite heterogeneity of treatment effect<br>• Pre-specify if you intend to conduct and report subgroup/stratified analysis based on the expected | effect modifiers with a pre-specified analysis strategy, these must be confirmed according to the statistical analyses plan: sub-group analysis; consider also comparing the effect size estimates obtained from the subgroups;<br>• If sub-group analysis are conducted but were not pre-specified in the statistical analyses plan, one must be very clear about *post-hoc* analysis;<br>• When no hypothesis was made on effect modifiers, this is still necessary to identifying heterogeneity of treatment effects: Outcome Weighted Learning (OWL) approaches [50] or other machine learning techniques [51] are ways to identify heterogeneity of treatment effect, and find covariates that demonstrate different treatment effect within subgroups defined by the covariates (these do not adjust for confounding but identify heterogeneity); |

This article is protected by copyright. All rights reserved.

| | | |
|---|---|---|
| | heterogeneity of treatment effect [49];<br>• Statistical power needs to be anticipated accordingly | |
| **1.4. Absence of a standard and strict definition of disease** | Generalisability issues:<br>• Decide whether physicians should be provided with a standard definition of the disease of interest, and clarify eligibility criteria – although this is a less pragmatic approach<br>• Review the databases of all sites before patient recruitment to identify the criteria used to define the disease of interest compared;<br><br>Statistical power:<br>• Anticipate the need to increase sample size to address the increased variance;<br>• Consider enrichment designs, e.g., the recruitment of patient sub-groups e.g., based on disease severity; when sample size is reached for one group stop recruiting and continue for other sub-groups until sample size is reached (i.e., to gain adequate power for each sub-group analysis) | Use random-effects model including random effects per disease severity subgroup to account for heterogeneity across these subgroups |
| **1.5. Motivation to take part in the trial and generalizability of results** | What is the target "routine clinical care" population for the trial?<br>• When designing the trial and using other sources of data (e.g., electronic medical records, claims databases, disease registries), describe the "routine clinical care" population, characteristics' distribution, in order to understand what is the routine clinical care population, and try to recruit a representative sample; this population may be different from this of traditional RCTs;<br>• Review the databases of all | Identify a possible self-selection bias<br>• The participants included in the PCT can be compared to the target "routine clinical care" population (cohort studies, registries etc.) in order to identify any observed differences; note that this comparison does not provide information on whether these differences will interact with the treatment effects, which would cause an error in the treatment average effect;<br>• Describe the characteristics of eligible subjects who do or do not agree to participate; describe |

Accepted Article

This article is protected by copyright. All rights reserved.

| | | |
|---|---|---|
| | eligible sites before trial kick-off to identify the adequate criteria to be used so as to recruit the "routine clinical care" population;<br><br>Adapt the design<br>• To minimize self-selection and improve the representativeness of the trial population, an opt-out consent framework may be considered [52].<br>• Consider exclusion criteria consistent with real world acceptable clinical use (e.g., drug use & safety issues).<br>• Consider cluster randomisation (randomising sites to ensure that all patients and treatment varieties are included<br><br>Identify self-selection<br>• The barriers to participation and patterns of self-selection may be explored through qualitative and quantitative evaluation<br>• When allowed, plan to collect key characteristics and reasons for not participating of eligible subjects who decline participation | reasons for not participating;<br><br>The analyses depend on the trial estimand:<br>• If the trial aims to measure the treatment effect in the trial population, then no analyses is necessary to correct for a selection bias;<br>• If the trial aims to measure the treatment effect in the target "routine clinical care" population and the trial populations differ from it, then techniques such as re-weighting [53] or Bayesian techniques may be considered to predict the effectiveness in a more representative population; this must be pre-defined in the analysis plan; |
| **2. Recruiting sites** | | |
| **2.1. Heterogeneity of care provided by participating sites** | Participating sites should reflect the actual heterogeneity observed in the real world regarding sites, e.g., the quality of medical care, the availability of staff, the type of patients, or treatments prescribed.<br><br>To understand and maximize heterogeneity:<br>• A framework can be built for understanding how to be "representative"<br>• From a pool of eligible sites, consider selecting sites through stratified sampling (e.g., oversampling of sites that are | Exploration phase:<br>• Explore if site characteristics had any impact on how patients were managed during the trial; this helps understanding what treatment effect comes from the "standard of care" and what effect comes from the intervention itself; adjustment of analyses may be necessary if important differences appear between sites;<br><br>Adjustment methods:<br>• Site-level adjustment is specifically relevant if local guidelines had an impact on how |

Accepted Article

This article is protected by copyright. All rights reserved.

Accepted Article

| | | |
|---|---|---|
| | underrepresented in the pool of eligible sites) to enhance representativeness<br>• In sites that are not used to deliver the intervention of interest, pre-specify a time period that allows all sites to go through a "learning phase"<br>• If possible, use two active treatments common for all sites, and have "usual care" (as defined by each site) as third arm<br><br>Identify a selection bias<br>Plan to collect key site-level data of participating sites, e.g., number of physicians, number of patients, equipment, specific clinical and/or patient management guidelines; | patients were managed;<br>• Account for inter-site differences in the analysis: random-effect models (with a random intercept at a site level); note this is appropriate for both patient-level trials and CRTs; weighted analyses to adjust for the over or under-representation of more "experienced" sites for instance |
| **2.2. Heterogeneity in the availability of treatment strategies** | This issue needs to be anticipated:<br>• Eligible sites that cannot deliver the intervention/comparator of the trial should not participate;<br>• If all sites are not able to provide the same treatment strategies (e.g., A, B & C), pre-specify if you will allow these sites to participate in the trial, before site recruitment;<br><br>Have several arms to the trial:<br>• If the comparator is "treatment as usual", it is important that each site provides a definition of "treatment as usual", and to plan for a 3-arm trial: intervention / comparator / and "treatment as usual";<br>• A "menu-driven" trial may be considered (sites select which treatment regimen they can choose as the comparator they are used to; e.g., the protocol includes the option for treatments A, B, C, & D; if a site only selects the interventions C & D, when comparing all treatment effects these sites are the | Consider network meta-analysis (NMA), preferably using individual patient data, to estimate effects of the different treatment strategies. Inter-site differences between similar treatment strategies could be accounted for by means of random-effects modelling, provided sufficient numbers of sites provide similar treatment strategies |

This article is protected by copyright. All rights reserved.

| | | |
|---|---|---|
| | excluded from the analysis); | |
| **2.3. Heterogeneity of data measurement and collection** | What is good quality data collection?<br>• Pre-define standard procedures to describe the expected data and the processes to check data once collected, e.g., plans to investigate outliers in the study data and plans for screening of data prior to analysis;<br>• Include a "data quality plan" to ensure good quality of key data in each site;<br><br>Assess sites prior to recruitment, in terms of<br>• Robustness and quality of data collection tools and systems;<br>• Data completeness and accuracy: use historic data from sites to ensure the data to be collected in the PCT can be reliably ascertained from the site database (i.e., if electronic health records are being used);<br>• Check the data system used to collect medical data, and in particular if the outcome of interest can be collected; if not, are there alternative medical data that could be used as a proxy for the primary endpoint?<br><br>From this evaluation:<br>• Identify practices that are "outliers" that may be excluded prior to recruitment;<br>• Identify what support might be needed by recruiting sites; provide training to sites on tools that will be used and monitor their implementation;<br>• Use statistical tools to identify low-performing site and plan corrective actions (e.g., site with poor coding vs average site, or site with high percentage of missing data etc.). | The issue of data heterogeneity across sites in terms of data collection quality cannot be tackled in the data analyses. |

Accepted Article

This article is protected by copyright. All rights reserved.

| | | |
|---|---|---|
| | Of note: audits and checks in sites can be misleading, i.e., using data clarification form (when asking a site to complete CRF, often the eHR is copied into the CRF and no additional information is gained). One can overcome this issue by using a linked data source to check the quality of the data, or measure the outcome in a different way (e.g., directly asking patients). | |
| **2.4. Heterogeneity in the ability to take part in the trial trial** | Check that eligible sites can be compliant with the trial protocol:<br>• Emphasize the availability and feasibility of good data collection<br>• Describe the recruitment process of sites, if known;<br>• If the process of recruitment is unknown, eHRs may be used to understand how participants are screened and recruited, compared to all patients who may be eligible<br><br>Collect information on eligible sites that do and do not participate to explore for a possible selection bias. | Compare sites that did and did not participate, e.g., relative to their experience of clinical research, the process of recruitment may not be random<br><br>If there is an over-representation of "more experienced" sites, then weighted analysis/reweighing methods may be performed |
| **2.5. Fragmentation of care** | Assess the risk of missing outcome data prior to recruitment:<br>• Robustness and quality of data collection tools and systems;<br>• Data completeness and accuracy<br>• If eHRs are planned to be used, run prior checks to ensure that the data to be collected in the PCT can be reliably ascertained from the site database;<br><br>Minimize the risk of missing outcome data<br>• Collect data through different data sources (e.g., eHRs, asking the patient etc.) and use linked data to capture all relevant information. | Management of missing outcome data: see item 5.2 below. |

Accepted Article

This article is protected by copyright. All rights reserved.

| **Randomisation, blinding and intervention** | | |
|---|---|---|
| **3.1. Absence of blinding and patient expectations** | Anticipate the issue of contamination by trial design: consider using a cluster randomized trial (CRT) if contamination is expected to have substantial impact<br><br>Choose an adequate endpoint to minimize subjectivity<br>• Use hard primary endpoints if possible to minimize subjectivity bias (e.g., hospitalization) [3, 8]; however, the use of hard endpoints does not entirely solve the problem because "knowledge" of the treatment may (differentially) affect other health behaviour or concomitant drug use in patients; to be able to account for this confounding effect, information on other health behaviour and concomitant drugs need to be collected;<br>Note: when a PROM is used the impact of subjectivity is even more problematic;<br><br>Plan to collect information and to measure<br>• Patients' expectation, e.g., which treatment the patient wants to receive at baseline: important aspect of PCTs because in real world, this will play a role in how the treatment "works";<br>• Drop-out rates (and where possible, the reason for drop-out);<br><br>Regarding outcome measurement<br>• Outcome measurement may be adjudicated by blinded medical experts; in addition, it is useful to blind the outcome assessor. | The potential impact of the absence of blinding needs to be described including a bias assessment regarding the processes for outcome data collection<br><br>Adjust the analyses, e.g., on compliance to protocol (drop-out) as well as health behaviours.<br><br>See also item 3.2 below. |
| **3.2. Treatment switching,** | Plan to collect information and to measure | The analytic approach needs to be aligned with the estimand: |

This article is protected by copyright. All rights reserved.

| | | |
|---|---|---|
| **cross-over between treatment arms** | treatment switching using pre-established data sources and tools: reason for switching, switching date, severity of disease prior switching, etc.;<br><br>Minimize the risk of switching if this would be a problem for data analyses:<br>• Inform physicians participants that for the period of the trial treatment switching should be avoided;<br>• In case there is a risk of substantial proportion of treatment switching, physicians' preference, or subversion, the Zelen design can be considered [54]; this design intends to facilitate clinicians' and patients' participation by randomizing patients prior to participation consent.<br><br>Treatment switching or discontinuation may also be used as pragmatic primary endpoints [37];<br><br>Establish a time to assess the primary endpoint that is not too far after likely treatment switching to avoid treatment effects being confounded by post-switching therapies. | If the question is: "what is the impact of initiating the treatment?" then the primary objective is measure the effect of being assigned to / initiating a treatment (and not a sequence of interventions), [55]:<br>• An ITT approach should be adopted to measure the effectiveness of treatment;<br>• In addition, the effect of treatment that would have been observed in the absence of switching ("if on-treatment" effect) can be modelled using key information (e.g., reason for switching, association between switching and outcome, etc.): inverse probability of censoring weights or g-computation may be used to take account of time-varying confounders and get closer to causal inference.<br><br>If the primary objective is to measure the effect of the treatment actually received<br>• Per protocol analyses could be performed provided that appropriate adjustment for bias can made (same as above) [56]; |
| **3.3. Physicians' preference and subversion** | The non-inclusion of particular eligible patients needs to be avoided<br>• Ensure that the identification of eligible patients is done before randomisation for both patient-level and site-level randomised trials<br>• If prior identification is not possible, then an independent recruiter should recruit participants; | Is there a difference at baseline between participants of each intervention group?<br>• Baseline characteristics of participants from each intervention group should be described and compared to identify any selection bias and confounding due to the non-inclusion by physicians of particular patients.<br>• The Bergner-Exner test can be used to identify whether subversion has occurred when |

Accepted Article

This article is protected by copyright. All rights reserved.

Accepted Article

| | | |
|---|---|---|
| | Alternative types of randomisation can be used:<br><br>For cluster-randomized trials<br>• The covariate-constrained randomisation [19] may be used to achieve balanced study arms: collect baseline data at a cluster-level or at a patient-level, aggregated to identify confounding.<br>• Stratification aims at achieving equal (or nearly equal) distributions of certain cluster characteristics in each study arm;<br><br>For randomisation at the patient-level<br>• Pseudo-cluster randomisation may be used, decreasing the risk of subversion; contamination is reduced because only a minority of control-arm participants are treated by majority-intervention providers;<br>• Random block sizes or minimization with a random twist;<br>• Perform randomisation with a remote randomisation system in order to have appropriate concealment/randomisation;<br><br>In the absence of alternatives, consider using a stepped wedge design. | block randomisation by site was used;<br><br>What is the impact of this difference?<br>• The impact of a range of selection probabilities can be assessed, e.g., using probabilistic bias analysis (PBA) [57];<br><br>How to correct for imbalanced participants characteristics at baseline? Several approaches can be considered including:<br>• pair-matching analysis<br>• stratification, stratification and matching – the most commonly used strategies<br>• multivariate regression models; however, this is insufficient to obtain equivalent intervention group as they do not take account of unobserved confounding |
| **3.4. Selection bias in cohort-multiple RCTs being differential between the intervention and the control arm** | Minimize the risk of refusal to participate to a PCT:<br>• Upon the first stage of participation consent (consent for data collection), all patients can be informed that during the cohort study period some new treatments may become available; patients can be asked whether they would be interested in taking the new treatment when/if this becomes available; | Several approaches are possible including ITT, PP analyses, or the use of instrumental variables.<br>The type of analyses that can be used depends on whether refusal is correlated with the outcome or not [28, 38]. |

This article is protected by copyright. All rights reserved.

| | | |
|---|---|---|
| | • Patients that refuse the possibility of a new treatment would subsequently (if a PCT is conducted) not be eligible for the trial and for randomisation;<br><br>Adapt the sample size<br>• The required sample size can be updated during the trial as more information about the refusal rate is gained; | |
| **4. Outcome selection and measurement** | | |
| **4.1. Choice of a subjective endpoints** | To choose appropriate endpoints: When several possible endpoints are relevant to assess the treatment's effectiveness and include both hard and soft endpoints:<br>• The selection of several relevant endpoints should be justified by a known relationship between subjective and hard endpoints (this can be found through a literature search); the risk of discrepant results needs to be anticipated;<br>• For hard endpoints, international standard definitions used by physicians in the real world should be used (e.g., death, progression of cancer);<br><br>To minimize the risk of subjectivity:<br>• The accessor/analyst can be blinded to avoid classification bias on outcome<br>• The outcome can be adjudicated by blinded medical experts<br><br>To explore the effect of subjectivity<br>• Collect patients' preference on treatments at baseline (by asking patients) | Explore whether the patients' preference modifies the association between treatment arm and outcome (interaction test by patients' preference);<br><br>Estimate how specific and sensible the subjective endpoints are:<br>• Some analyses can be conducted in a sub-sample of patients to measure the sensitivity and specificity of endpoints that are prone to measurement errors (e.g., progression of symptoms vs. radiologic progression).<br>• If the endpoint is prone to substantial measurement errors, include this additional information in the analysis, e.g., through regression calibration, imputation, or other ways of dealing with measurement error. |
| **4.2. Difficulties in collecting** | The risk of missing outcome data has several possible sources: | The sources and extent of missing outcome data need to be described: |

This article is protected by copyright. All rights reserved.

| | | |
|---|---|---|
| **outcome data, missing outcome data** | attrition, site, and timing of medical data collection; various approaches are possible (see items 1.1., 2.3., 2.5. and 4.3).<br>In addition to these set up solutions, others are suggested<br><br>• A hybrid approach for data collection can be considered e.g., combining ad hoc eCRFs and routinely collected data with minimal interference with routine clinical practice;<br>• Data collection should be kept at the minimum required to address study objectives;<br>• Outcome assessment should be made as easy as possible | • Is this related to some sites?<br>• Is it systematically missing sometimes missing (e.g., unscheduled visits always missing for some or all sites;<br><br>Then, the type of missingness needs to be evaluated:<br>• Missing completely at random, MCAR<br>• Missing at random, MAR (the tendency for data to be missing is not related to the value of missing outcome data, but to some of the observed data)<br>• Not missing at random, NMAR (the tendency for data to be missing is related to the value of missing outcome data); |
| **4.3. Timing of outcome measurement** | Planning of data collection:<br>• Specify scheduled visits for certain outcomes<br>• Specify timing and frequency of medical data collection;<br>• Define regular time periods over which to collect data | Evaluate whether patients for whom the outcome is present are different (at baseline) from patients for whom the outcome is missing;<br><br>For those who do not have a measurement close to the scheduled recording of the event (if any schedule visit), take the latest value before and the first value after the scheduled event and average them (within some acceptable pre-specified time window);<br><br>Consider interpolation/extrapolation of time trends (interpolation: taking timing of measurements into account) to impute the missing data;<br><br>Allow variable measurement schedules (as long as dates or elapsed time since randomisation are recorded) and use modelling to assess time trends rather than forcing a particular measurement schedule;<br><br>If adjustment for medication switching is needed, obtain outcomes at the point of switching, |

Accepted Article

This article is protected by copyright. All rights reserved.

| | | |
|---|---|---|
| | | as a critical confounder for next treatment. |
| **5. Data analysis** | | |
| **5.1. Analysis aligned with the estimand** | Clearly specify in the protocol which effect/aspect of the intervention is the target of inference: what is/are the trial estimand/estimands?<br><br>• Refer to the ICH E9(R1) addendum [18] on how clinical trial protocols and statistical analysis plans should be written and implemented [58];<br>• Consider in the statistical analyses plan the analytical approaches that are aligned with the estimand;<br>• Specify if we are interested in making causal inference? [59] | See solutions detailed above in items 1.2 (Absence of adherence-enhancement strategy), 1.3. (Heterogeneity of treatment effects), 1.5. (Motivation to take part in the trial and generalizability of results) and 3.2 (Treatment switching). |
| **5.2. Missing outcome data** | In the statistical analysis plan:<br><br>• Assumptions on missingness can be anticipated; explicitly explain the assumptions made and methods used to handle missingness;<br>• Methods to manage missing outcome data – due to attrition, non-adherence to treatment, treatment switching, or any other reason – should be clearly specified;<br>• This method needs also to be chosen in the light of the estimand [60]; | Depending on missingness assumption, missing outcome data can be handled in various ways:<br>• Multiple imputation techniques<br>• Likelihood-based methods<br>• Dependent censoring approaches<br>• Adjusting the analysis for factors, propensity scores<br><br>Several approaches may be combined using sensitivity analyses [61] e.g., inverse probability of censoring weighted (IPCW), multiple imputation, compliance mixture models, time varying Cox models.<br><br>Sensitivity analyses may be performed using the worst-case and best-case scenario, to explore the consequences of imputing the outcome on treatment effect.<br><br>Complete-case analysis should be avoided as this assumes missing data |

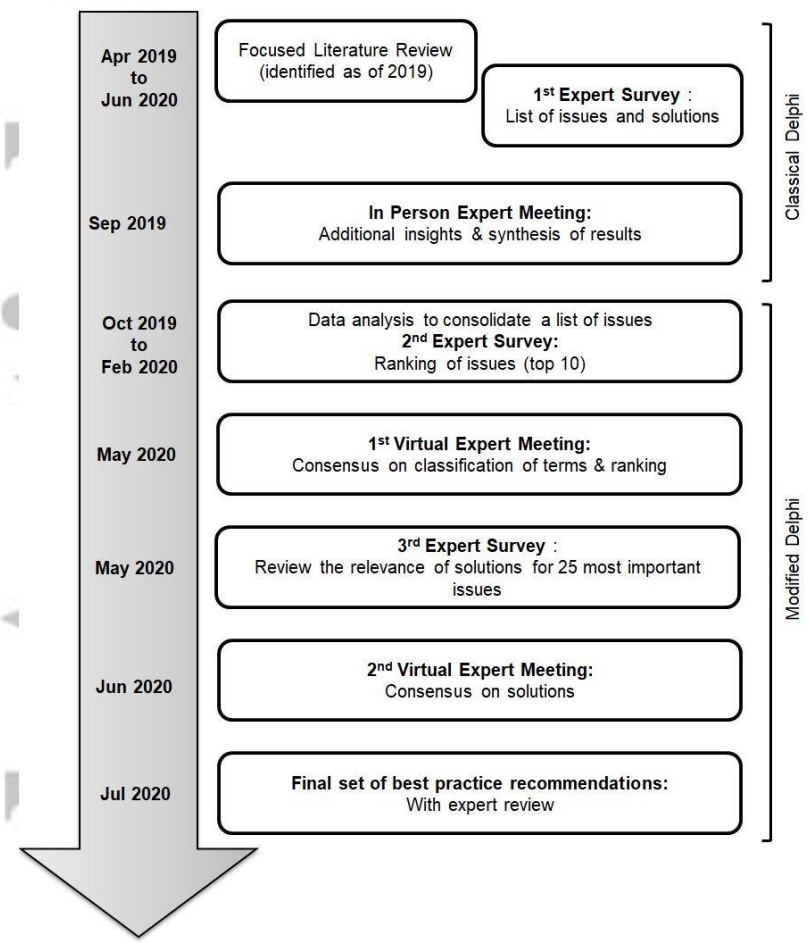
This article is protected by copyright. All rights reserved.

| | | |
|---|---|---|
| | | is not informative, which is usually not the case. |
| **5.3. Lack of statistical power** | Anticipate any loss of power by increasing the sample size.<br><br>Calculate sample sizes under a range of assumptions with respect to different baseline rates of the outcome. | Include informative priors in a Bayesian analysis. However, informative prior may be less acceptable to regulatory authorities.<br><br>Perform interim analyses (may lead to early termination of the trial because effectiveness is sufficiently established, or because of futility). |

Accepted Article

This article is protected by copyright. All rights reserved.

**Figure 1:** Iteration process of data generation and consensus building among experts




This article is protected by copyright. All rights reserved.