



Published in final edited form as:

J Clin Epidemiol. 2018 November ; 103: 10–21. doi:10.1016/j.jclinepi.2018.06.009.

Patients and investigators prefer measures of absolute risk in subgroups for pragmatic randomized trials

Eleanor J. Murray^a, Ellen C. Caniglia^a, Sonja A. Swanson^{a,b}, Sonia Hernández-Díaz^a, and Miguel A. Hernán^{a,c}

^aHarvard T. H. Chan School of Public Health, 677 Huntington Ave, Boston, MA, 02115

^bDepartment of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands ^cHarvard-MIT Division of Health Sciences and Technology, Boston, MA, 02139

Abstract

Objectives—Pragmatic randomized trials are important tools for shared decision-making, but no guidance exists on patients’ preferences for types of causal information. We aimed to assess preferences of patients and investigators towards causal effects in pragmatic randomized trials.

Study Design and Setting—We: (a) held 3 focus groups with patients (n=23) in Boston, MA; (b) surveyed (n=12) and interviewed (n=5) investigators with experience conducting pragmatic trials; and (c) conducted a systematic literature review of pragmatic trials (n=63).

Results—Patients were distrustful of new-to-market medications unless substantially more effective than existing choices, preferred stratified absolute risks, and valued adherence-adjusted analyses when they expected to adhere. Investigators wanted both intention-to-treat and per-protocol effects, but felt methods for estimating per-protocol effects were lacking. When estimating per-protocol effects, many pragmatic trials used inappropriate methods to adjust for adherence and loss to follow-up.

Conclusion—We make 4 recommendations for pragmatic trials to improve patient centeredness: (1) focus on superiority in effectiveness or safety, rather than noninferiority; (2) involve patients in specifying *a priori* subgroups; (3) report absolute measures of risk, and (4) complement intention-to-treat effect estimates with valid per-protocol effect estimates.

Keywords

pragmatic trial; per-protocol; adherence adjustment; intention-to-treat; patient preferences; health communication; causal inference

Corresponding author: Eleanor Murray, Harvard TH Chan School of Public Health, 677 Huntington Ave, Suite 911, Boston, MA, USA, 02115, emurray@mail.harvard.edu.

Disclaimer: All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology Committee.

1. Introduction

Patient involvement in making medical decisions, or shared decision-making, has been shown to be important for patient satisfaction, treatment adherence, and health outcomes,(1–3) and a large literature exists on methods for improving doctor-patient communication and shared decision-making.(3–7) However, neither preferences of patients and investigators nor current practices regarding choice of causal contrasts have been systematically characterized in pragmatic randomized trials, which are designed to address real-world questions about healthcare options.

For example, non-adherence and loss to follow-up in pragmatic trials compromises the interpretability of the usual intention-to-treat effect estimates, which may need to be complemented by other measures of causal effect, such as the per-protocol effect, i.e., the effect that would have been observed if patients and clinicians had fully adhered to the study protocol. The preferences of patients and investigators regarding per-protocol effects are largely unknown.

To help fill these knowledge gaps about preference of causal contrasts, we: (a) conducted focus groups with patients to determine their preferences; (b) interviewed and surveyed principal investigators of pragmatic trials to determine their preferences and their perceived barriers to estimating and reporting causal contrasts; and (c) conducted a systematic literature review of pragmatic trials published in major medical journals to describe current practices for conducting and reporting causal effects.

2. Methods

2.1. Patient focus groups

We conducted three focus groups of patients aged 18 years or older with a chronic medical condition requiring regular medication or physician visits, with no restrictions on type or duration of condition. Participants were recruited from neurology, psychology, gastroenterology, and renal outpatient clinics at Brigham & Women’s Hospital in Boston. Study pamphlets were placed throughout clinics and distributed through the Dana Farber Cancer Center social work group. A member of the study team (EM) was available to answer questions, assess eligibility, and enroll participants, or patients could enroll via phone or email. Patients were excluded if they could not sit for prolonged periods, could not make medical decisions due to a neurological condition, or were not available at scheduled group times (Figure 1).

Focus groups were comprised of 6–8 individuals, approximately 90 minutes long, and conducted in English. Patients were compensated with a gift card. An experienced researcher from the Harvard Derek Bok Center for Teaching and Learning (Dr. Jenny Bergeron) moderated the focus groups. Patients were presented three vignettes, based on real-world trials designed to assess patients’ preferences when deciding between medications with (a) non-adherence related to convenience; (b) non-adherence more common among people at higher risk for the outcome (heart attack); or (c) differing side

effect risks (Appendix A). Each session was transcribed verbatim by an independent contractor.

2.2. Investigator interviews

We conducted five one-on-one telephone interviews with a convenience sample of principal investigators of one or more pragmatic trials, identified through the authors' professional networks. Potential investigators were contacted by email; those who were unavailable or unwilling were asked to suggest an alternate investigator. All but one investigator, who recommended a colleague, agreed to participate. All investigators had an affiliation with Harvard University, and had previously collaborated with our research group. Four were based primarily in the Northeastern United States, and one in Europe.

The interviewer (EM) followed a semi-structured guide (Appendix B) and began by asking for a definition or description of pragmatic trials. Interviewees were led through a series of questions on their research and their most recent pragmatic trial, followed by a discussion of conducting and reporting pragmatic trials. Interviews were transcribed by the interviewer.

2.3. Investigator surveys

We conducted an online survey (Qualtrics, Provo, UT) of principal investigators who had received funding from the Patient Centered Outcomes Research Institute (PCORI) pragmatic trials mechanism (n = 24). Investigators were contacted by email up to 3 times and could enter a gift card raffle as incentive. The survey was conducted after the focus groups, interviews, and literature review, and designed to target themes identified from those. In total, 12 investigators completed the survey, which asked about features of pragmatic trials, trial experiences, and knowledge, attitudes and practices regarding pragmatic trial analyses. All questions had multiple pre-determined answer choices, but investigators could also provide free-form text answers (Appendix C).

2.4. Qualitative analysis

Patient focus groups and investigator interviews were coded from transcripts (EM). Survey responses were tabulated in the Qualtrics platform. Focus group codes were summarized within and across vignettes and focus groups, and interview codes were summarized within and across interviews. Major themes were identified from frequently occurring codes using Excel (EM), and cross-cutting themes emerging from focus groups and interviews were identified. Two additional study team members (EC, SS) reviewed all transcripts and resulting themes.

2.5. Literature review

We conducted a systematic literature review of pragmatic trials published in the New England Journal of Medicine (NEJM), the Journal of the American Medical Association (JAMA), the British Medical Journal (BMJ), and The Lancet within the past 10 years (September 1, 2006 to September 2, 2016). We selected these journals to be representative of perceived best-practices in pragmatic trial research among medical research community.

We searched the NCBI/PubMed database on Sept 6, 2016, with a search strategy combining terms for randomized trials with terms for pragmatic trial characteristics or phase IV trials (Appendix D). Randomized trial search terms were based on an existing validated search strategy(8) while search terms for pragmatic trials were identified based on a preliminary search for studies with ‘pragmatic trial’ in the title or abstract. Only trials of medical treatments and health-related outcomes were eligible. Phase I, II, and III trials were excluded by definition, as were cluster-randomized or cross-over studies, trial protocols, and articles reporting only secondary cost-effectiveness analyses.

We required trials be closer to the pragmatic extreme of the pragmatic-explanatory continuum described by Thorpe et al(9)(Appendix E). Briefly, design elements of randomized trials at the pragmatic extreme are more similar to clinical practice or routinely collected observational data than to laboratory-style highly controlled experiments. For example, a pragmatic trial might enroll all patients with a set of suggestive symptoms, whereas an explanatory trial could require a diagnosis and confirmatory biomarker or bloodwork results.

The screening process involved four stages (Figure 2). First, all articles underwent title, brief abstract, and trial register review to exclude ineligible study types. Eligible articles then underwent full abstract review by 2 of 3 independent reviewers (EM, EC, SS) using a standardized screening form (Appendix F). Abstracts were scored on a 5-point scale (1 = pragmatic, 3 = unclear, 5 = explanatory) for four characteristics--study population, interventions, blinding, and outcomes--based on a simplified PreciS-2 tool (Appendix F).(10) When reviewers disagreed, the third served as tie-breaker. Abstracts averaging 3 or less underwent full-text review scored on a 3-point scale (0 = pragmatic, 1 = intermediate, 2 = explanatory); articles averaging above 1 were excluded (Appendix G).

One of two reviewers (EM and EC) extracted information on design and analysis from each selected trial (Appendix H), and data extraction was checked by the other reviewer to ensure accuracy. We compared trial characteristics to major qualitative themes to assess concordance between patient and investigator preferences and current practices in reporting of pragmatic trials.

3. Characteristics of participants

3.1. Patients in focus groups

In total, 23 patients attended a focus group (Figure 1, Table 1). Attendees were 74% female, 39% African American, and 17% Hispanic. The mean age was 56 years (range 23 to 82 years). 61% of participants had a college degree, while 26% had no college education, and 9% had not completed high-school. 10 patients completed eligibility screening but were unavailable or did not attend a focus group. Men, patients younger than 35 years old, or individuals with ‘some college’ were somewhat more likely not to attend (Appendix I).

3.2 Investigators

The 5 interviewed investigators reported a combined 83 years of experience conducting trials and on average had been involved in 3 pragmatic trials (Table 1). All were currently

involved in at least one pragmatic trial, ranging in size from 300 to 180,000 (cluster-randomized) participants, and duration from 12 months to 20 years.

The 12 survey respondents reported a combined 100 years of experience conducting trials, had on average been involved in 3 pragmatic trials, and were currently involved in at least one pragmatic trial. Trials ranged from 900 to 90,000 participants (one investigator reported a cluster-randomized trial of unspecified size). Trial duration ranged from 12 months to 3 years.

Pragmatic trials in literature review—We identified 287 potential articles, of which 75 had full-text review (Figure 2). Sixty-eight articles reporting on 63 trials met our final eligibility criteria (Appendix G, Appendix H). Of 5 trials with multiple articles, 2 articles reported extended follow-up and were excluded. (11, 12) The other 3 had paired trial designs, with 2 sets of eligibility criteria and stratified randomization; for the review, we pooled paired trials.

Table 2 describes included trials, which ranged from 60 to over 31,000 participants (median: 511; IQR: 241, 822), and from 12 hours to 15 years (median: 12 months; IQR: 6, 24 months). The most common interventions were medications (n=18 trials), surgery (n=13), and medical devices (n=12). 38 trials assessed patient-centered outcomes (symptom severity or resolution). All but three (13–15) assessed interventions and/or comparators available outside the trial; only two studies blinded participants and/or investigators. (16, 17)

Strategies allowing patients with strong preferences or indications to forego randomization were uncommon: 1 trial included patient-preference arms, and 1 physician-preference arms. (18, 19) One vaccination trial allowed optional re-randomization of patients in the second year. (16)

4. Summary of Findings

The key themes, identified across focus groups, interviews, and the literature review, are summarized below, with representative patient and investigator quotations for each theme in Table 3.

4.1. Characterizing pragmatic trials

We adopted a standard definition of pragmatic trials in the patient focus groups (implicitly) and the systematic review (explicitly). We asked interviewed investigators to provide their own definitions and used these to design a survey question on key features of pragmatic trials.

In the interviews, investigators focused on design features of eligibility criteria, study sites, and interventions as key pragmatic trial characteristics. They stressed that these features increased generalizability to clinically-relevant populations and health care settings and led to more realistic comparisons. Pragmatic trials were described as important for assessing feasibility of intervention implementation, and assessing methods of promoting adherence.

No investigators spontaneously discussed post-randomization features, such as increased non-adherence or loss to follow-up.

When asked to select key characteristics differentiating pragmatic trials from randomized clinical trials, surveyed investigators chose ‘more representative study sample’ (n=11 of 12), ‘looser inclusion/exclusion criteria’ (n=9), ‘use of usual care comparator treatment/intervention’ (n=6), ‘potential for non-adherence’ (n=5), and ‘use of patient-reported or patient-centered outcomes’ (n=4). Two investigators wrote about concerns for increased potential for variable administration of the intervention depending on whether administered by clinicians or patients for example, or because of lower levels of provider expertise or training.

4.2. Superiority vs non-inferiority

Patients displayed a strong preference towards established treatments. When asked to choose between medications with similar benefits and harms, nearly half preferred the standard medication, and only 1 in 10, the new medication. The remainder either had no preference (about 1 in 5), or requested additional information, such as side effects, medication interactions, and trial participants’ use of non-medical alternatives (diet, exercise, acupuncture) which might impact trial estimates. Patients described standard medications as ‘tried and true’ and felt there was no reason to switch to a new medication which was as effective as an existing one. Some even suggested that doctors recommending a new, equally effective, medication must be receiving a pay-out or other benefit. These responses suggest that patients value superiority over non-inferiority contrasts.

All interviewed investigators reported interest in effectiveness and superiority of the intervention as primary goals of their current trials, and 9 of 12 surveyed investigators were currently conducting a pragmatic trial with a superiority goal, compared to 3 with a non-inferiority goal. Objections to non-inferiority included incompatibility with the aims of pragmatic trials, need for good working knowledge of how to approach uptake of an intervention in a population (if that knowledge existed, the pragmatic trial would be unnecessary), and challenges for estimating power.

Consistent with stated preferences of patients and investigators, 45 of 63 pragmatic trials in the literature review assessed superiority for effectiveness alone; 9 assessed superiority for both effectiveness and safety; and 2 both superiority and non-inferiority for effectiveness (Table 2). (20, 21)

4.3. Risk-benefit profile

We assessed patient preferences regarding risk-benefit. When patients were presented with a choice between two medications with similar efficacy but different side-effect risks, almost all chose the standard medication with lower risk when the side effect was liver damage. About two-thirds were willing to try a new ‘slightly more effective’ medication with higher side effect risk when the side effect was moderate weight gain. Patients raised questions about potential interactions between the new medication and their current medications, concerned that new-to-market medications may be poorly understood in terms of safety compared to medications with longer track records of use.

Only one interviewed and one surveyed investigator planned *a priori* to assess safety in their current trial, but investigators believed that pragmatic trials can study risk-benefit profiles (that is, both effectiveness and safety outcomes). However, investigators suggested studying safety outcomes requires adjustment for adherence (see 4.5 below). One described a willingness to ‘give up a bit of randomization’ in exchange for a fuller understanding of safety implications.

Interestingly, despite patients’ and investigators’ interest in risk-benefit profiles, safety outcomes were uncommon in our review—only 10 of the 63 trials reported safety analyses in addition to effectiveness analyses (Table 2), and none reported safety analyses alone.

4.5. Adherence adjustment: Intention-to-treat vs per-protocol effects

The type of information patients relied on for decision-making was highly context-dependent. While patients appreciated the value of estimating the overall intention-to-treat effect, they preferred subgroup intention-to-treat analyses based on more personalized information. When trial adherence was discussed but reasons for non-adherence were not given, patients relied on intention-to-treat information, but asked for details about those reasons, and about sub-group effects. However, when information on reasons for non-adherence were provided, some patients preferred the per-protocol effects. For example, when asked to choose between medications for which differences in convenience drove adherence, patients initially preferred the more convenient medication. However, if told taking the less convenient medication as prescribed would have led to fewer adverse events, some patients changed their decision. When the *standard* medication was inconvenient but safer, this ‘per-protocol’ argument led almost all patients to prefer the standard medication. In contrast, when the *new* medication was the inconvenient but safer option, the per-protocol argument was persuasive only for those who felt the level of inconvenience would not be a barrier to their own adherence (Table 3, theme 4).

Investigators stressed the importance of the intention-to-treat effect because of ‘the causal question it answers’ (n=9 of 12), ‘this effect is always estimated/industry standard’ (n=6), and ‘the ability to estimate it without bias’ (n=5) (Table 4). All surveyed investigators preferred intention-to-treat effects to evaluate effectiveness, whereas only 8 preferred it for safety – 3 now preferred per-protocol effects. All interviewed investigators, and 7 of 12 surveyed investigators, had also specified *a priori* adherence-adjusted analyses of primary or secondary outcomes in their own trials. Among the 7 surveyed investigators who planned per-protocol effect estimation, the most common reason was ‘the causal question it answers’ (Table 4). The main drivers behind interviewed investigators’ interest in adherence-adjusted effects were a desire to assess the ‘real’ effect of the intervention, an acknowledgement that adherence varies over time, and a perception of this effect being more generalizable. However, one-quarter of surveyed investigators were unfamiliar with methods for adjusting for post-randomization predictors of adherence (or loss to follow-up), and familiarity was correlated with intention to estimate per-protocol effects (Table 4).

In the literature review only approximately one-third (n=22) of the trials conducted some form of per-protocol analysis, and the most common approach (n=20) was a naïve ‘per-protocol population’ analysis, restricted to adherers with no adjustment for potential bias

caused by this restriction (Table 5). There was a lack of consensus on how to construct this per-protocol population, and whether to exclude individuals who were improperly randomized, experienced competing events (e.g. death, for symptom reduction outcomes), or were lost to follow-up; and several trials defined multiple per-protocol populations as sensitivity analyses. Only one trial used another method: adjustment for treatment received in a model comparing assignment and outcome.(18, 22) Finally, one trial conducted a landmark analysis with the goal of estimating the treatment effect in responders.(21)

Relatedly, of 55 trials without time-to-event outcomes, loss to follow-up adjustment was uncommon: 31 used complete case analysis only (Table 5); multiple imputation was used in only 11 trials (7 primary, 4 sensitivity analyses). Single-value imputation was also relatively common in primary (n=4) or secondary (n=2) analyses, imputing missing outcomes as events ('worst case') or non-events ('best case').

4.6. Absolute vs. relative measures of risks

Our focus group vignettes were intended to compare the importance of absolute and relative measures of effect by holding the relative measure constant while varying the absolute effect size. Patients were responsive to variations in absolute risk, but primarily based their decisions on the magnitude of the largest risk rather than differences or ratios. For example, a 3 in 1,000 risk of liver damage was unacceptable and 96% favored the standard, lower-risk, medication. In contrast, with a 3 in 1,000,000 risk of liver damage, 30% were willing to try the new, higher-risk, medication. Overall, few patients appeared to consider the lower risk information, except when risks were high for both medications: a fifth of patients requested a third option when liver damage risks were 3 in 10 versus 1 in 10 (Table 6).

Interviewed investigators all agreed on the importance of reporting absolute measures but were split on whether to report relative measures: 3 preferred reporting both, and 2 preferred absolute only (surveyed investigators were not asked about measures). Perceived benefits of absolute measures were clarity of information conveyed (especially in safety trials), ease of comparing burden of disease, and utility for setting priorities. In contrast, relative measures were perceived as more useful for extrapolation. Investigators also suggested using visuals (e.g. colored grids indicating risk), highlighted challenges of conveying information in high-emotion contexts (e.g., patients near death or with life-threatening conditions), and disagreed about how to convey results to clinicians or policy makers (2 felt only absolute measures were appropriate). One investigator noted "*for advocacy, as a tool, sometimes relative measures are better. If we found 60% improvement [for example], for advocacy we might want to say that instead of 3% versus 4%... but as a scientist, absolute measures are the most honest*" (Table 3).

Of the 63 trials, 43 reported absolute effect measures, either alone (n=25) or in addition to relative measures (n=17). Nearly all trials reported absolute occurrence measures, but 2 provided these only graphically.(23, 24) Finally, 17 trials included only relative measures, while a further 4 reported p-values for group comparisons but no effect estimates, only occurrence measures (Table 5).

5. Discussion and recommendations

Our patient focus groups, investigator interviews and surveys, and systematic literature review together suggest several recommendations for design, analysis, and reporting of pragmatic randomized trials, and identified methodological challenges for which improved guidance is needed.

5.1. Recommendation 1: Assess superiority rather than non-inferiority when the goal is improving shared decision-making

Patients were skeptical when offered a new medication of equal effectiveness to an existing medication, investigators perceived identifying the best treatment option as a key purpose of pragmatic trials, and roughly 9 in 10 trials in our review included a superiority contrast. There may be other reasons for assessing non-inferiority in randomized trials that were not assessed in our study, such as cost-effectiveness or safety-effectiveness trade-offs, but these should be justified or supplemented with superiority comparisons when improved patient involvement in decision-making is a trial goal.

5.2. Recommendation 2: Per-protocol effects should be reported with reasons for non-adherence to increase interpretability

Patients and investigators valued the intention-to-treat principle but expressed interest in adherence-adjusted effects for at least some scenarios, and many published trials presented adherence-adjusted estimates. Per-protocol effects were deemed most useful when patients expected themselves to be adherent or when trust in the medication was high. Information on reasons for non-adherence observed in pragmatic trials is therefore essential in helping patients assess their own expected adherence level, interpret per-protocol effects, and make informed treatment decisions. Trials should report the main reasons for non-adherence, not only the percent non-adherence.

5.3. Recommendation 3: Report absolute effect and occurrence measures with a combination of tabular and graphical results for increased comprehension

Patients were most comfortable assessing absolute outcomes, and investigators suggested absolute measures of effect were generally preferred for all audiences. Although investigators also felt visual tools were needed to help patients understand quantitative information, patients in our focus groups were nuanced and thoughtful when confronted with absolute risks of varying sizes (Table 6). Investigators were also concerned that clinicians and policy makers vary in their statistical literacy, and importantly, may be susceptible to being given biased perceptions of results based on the type of measure presented. Finally, nearly a fifth of trials provided only p-values and no point estimate of the effect. Although we did not ask patients or investigators about p-values, current statistical best practices discourage such reliance on p-values.⁽²⁵⁾ These findings are supported by literature on patient-clinician communication and statistical literacy.^(3, 26, 27) To ensure maximum comprehension, pragmatic trials should provide graphical or visual summaries as supplements to, rather than substitutes for, tabular results.⁽³⁾

5.4. Recommendation 4: Pragmatic trials should involve patients in pre-specifying subgroup analyses

Patients were interested in personalized decision-making and spontaneously requested results stratified by a wide range of characteristics, including demographics; cultural backgrounds; comorbidities and disease severity; lifestyle characteristics and co-behaviors, such as smoking, diet, or use of home remedies; and other medication usage. A broadly generalizable population was identified as a key characteristic of pragmatic trials by investigators, but only 16% of pragmatic trials presented stratified results. The most common stratification variables in these trials were age, gender, condition subtype, and markers of severity. Including patients in the design of pragmatic trials may help improve the relevance of *a priori* subgroup analyses for shared decision-making.

5.5. Methodologic challenges

Although per-protocol effect estimates appear to be of interest, clear guidance on appropriate adherence-adjustment methods is needed. Investigators felt that new methodology was needed and identified several key challenges such as uncertainty regarding confounders for adherence-adjusted effects, and large sample sizes required by some existing approaches. Trials included in our review often used methods which can introduce bias,(28) and survey results suggest this reflects a lack of awareness or familiarity with existing methods. Estimating per-protocol effects requires adjusting for post-randomization adherence and, therefore can introduce bias.(29) Adjusting for this bias is possible but requires controlling for post-randomization confounders of adherence and the outcome, using a method such as inverse probability weighting.(28–31)

Loss to follow-up was also a common problem in published trials, although not raised in our interviews. Trials generally relied on strong assumptions, such as non-informative censoring, or did not account for joint predictors of loss to follow-up and the outcome.(28, 32) Our finding mirrors a review of extended follow-up studies of trial populations, in which 51% of trials used a complete case approach, and only 25% included sensitivity analyses for missing data.(33) Improved guidance for trialists on the use of existing methods of loss to follow-up adjustment appears warranted.

5.6. Limitations

As with all qualitative studies, patients and investigators included in our study may not be representative of all patients or trialists. All patients lived in New England, and several reported participating in prior research studies. Although we were unable to ask patients about medical conditions or health insurance status, it is likely that most had some health insurance. Similarly, all interviewed investigators had prior collaborations with our research group and may therefore be better informed about causal inference definitions and methods than other trialists. However, focus group and interview results aligned well with the investigator survey responses and published trials, suggesting that the perspectives were not uncommon. Our study may therefore provide insight into patient preferences, and beliefs and practices of academic investigators interested in design issues for pragmatic trials, as well as help identify areas where further assessment of knowledge, attitudes, and practices would be beneficial.

5.7. Conclusions

Patients, investigators, and current practices in published literature suggest several considerations for the design and analysis of pragmatic randomized trials that could improve the utility of trial results for patient decision-making and to inform future research directions. Although our findings are limited by the small samples, patients and investigators strongly preferred superiority outcomes, and absolute measures of trial outcomes. Patients wanted detailed information about baseline characteristics, stratified treatment effects, and reasons for non-adherence to better evaluate the applicability of trial results to their own experiences. Investigators and trialists were interested in supplementing intention-to-treat effect estimates with per-protocol effect estimates, but need guidance on appropriate methods to adjust for non-adherence and loss to follow-up.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: This study was supported through a Patient Centered Outcomes Research Institute (PCORI) award (ME-1503-8119) and grant number T32 AI007433 from the National Institute of Allergy and Infectious Disease. All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology Committee. The funder approved the protocol but had no role in the conduct, analysis, or reporting of study findings.

The authors would like to thank the patients and investigators who participated in this study for sharing their time and experiences. The authors would also like to thank Jenny Bergeron, Courtney Hall, and Jeffery Solomon at the Derek Bok Center for Teaching and Learning, Harvard University School of Education, for conducting the focus groups (JB, CH), for providing feedback on study materials (JB, CH, JS), and Scott Lapinski at the Francis A. Countway Library of Medicine, Harvard Medical School, for his advice on developing a systematic search hedge.

Ethics approval: The focus groups and interviews were deemed exempt by the Harvard T.H. Chan Institutional Review Board (IRB); all participants gave informed consent before taking part in a focus group or interview.

References

1. Härter M, Simon D. Shared decision-making in diverse health care systems—Translating research into practice. *Patient Education and Counseling*. 2008; 73(3):399–401.
2. Joosten EAG, DeFuentes-Merillas L, de Weert GH, Sensky T, van der Staak CPF, de Jong CAJ. Systematic Review of the Effects of Shared Decision-Making on Patient Satisfaction, Treatment Adherence and Health Status. *Psychotherapy and psychosomatics*. 2008; 77(4):219–26. [PubMed: 18418028]
3. Zipkin DA, Umscheid CA, Keating NL, Allen E, Aung K, Beyth R, et al. Evidence-based risk communication: a systematic review. *Ann Intern Med*. 2014; 161(4):270–80. [PubMed: 25133362]
4. Epstein AM, Taylor WC, Seage GR 3rd. Effects of patients' socioeconomic status and physicians' training and practice on patient-doctor communication. *Am J Med*. 1985; 78(1):101–6. [PubMed: 3966475]
5. Holzel LP, Kriston L, Harter M. Patient preference for involvement, experienced involvement, decisional conflict, and satisfaction with physician: a structural equation model test. *BMC Health Serv Res*. 2013; 13:231. [PubMed: 23800366]
6. Mansell D, Poses RM, Kazis L, Duefield CA. Clinical factors that influence patients' desire for participation in decisions about illness. *Archives of Internal Medicine*. 2000; 160(19):2991–6. [PubMed: 11041908]

7. Francis V, Korsch BM, Morris MJ. Gaps in doctor-patient communication. Patients' response to medical advice. *N Engl J Med*. 1969; 280(10):535–40. [PubMed: 5764453]
8. McKibbin KA, Wilczynski NL, Haynes RB. Hedges Team. Retrieving randomized controlled trials from medline: a comparison of 38 published search filters. *Health Info Libr J*. 2009; 26(3):187–202. [PubMed: 19712211]
9. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol*. 2009; 62(5):464–75. [PubMed: 19348971]
10. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ: British Medical Journal*. 2015:350.
11. Grant AM, Cotton SC, Boachie C, Ramsay CR, Krukowski ZH, Heading RC, et al. Minimal access surgery compared with medical management for gastro-oesophageal reflux disease: five year follow-up of a randomised controlled trial (REFLUX). *BMJ (Clinical research ed)*. 2013; 346:f1908.
12. Abalos E, Addo V, Brocklehurst P, El Sheikh M, Farrell B, Gray S, et al. Caesarean section surgical techniques: 3 year follow-up of the CORONIS fractional, factorial, unmasked, randomised controlled trial. *Lancet (London, England)*. 2016; 388(10039):62–72.
13. Bullen C, Howe C, Laugesen M, McRobbie H, Parag V, Williman J, et al. Electronic cigarettes for smoking cessation: a randomised controlled trial. *Lancet (London, England)*. 2013; 382(9905): 1629–37.
14. Ikramuddin S, Korner J, Lee WJ, Connett JE, Inabnet WB, Billington CJ, et al. Roux-en-Y gastric bypass vs intensive medical management for the control of type 2 diabetes, hypertension, and hyperlipidemia: the Diabetes Surgery Study randomized clinical trial. *Jama*. 2013; 309(21):2240–9. [PubMed: 23736733]
15. Mouncey PR, Osborn TM, Power GS, Harrison DA, Sadique MZ, Grieve RD, et al. Trial of early, goal-directed resuscitation for septic shock. *The New England journal of medicine*. 2015; 372(14): 1301–11. [PubMed: 25776532]
16. DiazGranados CA, Dunning AJ, Kimmel M, Kirby D, Treanor J, Collins A, et al. Efficacy of high-dose versus standard-dose influenza vaccine in older adults. *The New England journal of medicine*. 2014; 371(7):635–45. [PubMed: 25119609]
17. Gabay C, Emery P, van Vollenhoven R, Dikranian A, Alten R, Pavelka K, et al. Tocilizumab monotherapy versus adalimumab monotherapy for treatment of rheumatoid arthritis (ADACTA): a randomised, double-blind, controlled phase 4 trial. *Lancet (London, England)*. 2013; 381(9877): 1541–50.
18. Grant AM, Wileman SM, Ramsay CR, Mowat NA, Krukowski ZH, Heading RC, et al. Minimal access surgery compared with medical management for chronic gastro-oesophageal reflux disease: UK collaborative randomised trial. *BMJ (Clinical research ed)*. 2008; 337:a2664.
19. Little P, Moore M, Kelly J, Williamson I, Leydon G, McDermott L, et al. Delayed antibiotic prescribing strategies for respiratory tract infections in primary care: pragmatic, factorial, randomised controlled trial. *BMJ (Clinical research ed)*. 2014; 348:g1606.
20. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, et al. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. *Bmj*. 2015; 351:h5627. [PubMed: 26559241]
21. Kirchhof P, Andresen D, Bosch R, Borggrefe M, Meinertz T, Parade U, et al. Short-term versus long-term antiarrhythmic drug treatment after cardioversion of atrial fibrillation (Flec-SL): a prospective, randomised, open-label, blinded endpoint assessment trial. *Lancet*. 2012; 380(9838): 238–46. [PubMed: 22713626]
22. Nagelkerke N, Fidler V, Bernsen R, Borgdorff M. Estimating treatment effects in randomized clinical trials in the presence of non-compliance. *Stat Med*. 2000; 19(14):1849–64. [PubMed: 10867675]
23. Gray R, Ives N, Rick C, Patel S, Gray A, Jenkinson C, et al. Long-term effectiveness of dopamine agonists and monoamine oxidase B inhibitors compared with levodopa as initial treatment for Parkinson's disease (PD MED): a large, open-label, pragmatic randomised trial. *Lancet*. 2014; 384(9949):1196–205. [PubMed: 24928805]

24. Lamb SE, Marsh JL, Hutton JL, Nakash R, Cooke MW. Mechanical supports for acute, severe ankle sprain: a pragmatic, multicentre, randomised controlled trial. *Lancet (London, England)*. 2009; 373(9663):575–81.
25. Wasserstein RL, Lazar NA. The ASA’s Statement on p-Values: Context, Process, and Purpose. *The American Statistician*. 2016; 70(2):129–33.
26. Allen M, MacLeod T, Handfield-Jones R, Sinclair D, Fleming M. Presentation of evidence in continuing medical education programs: a mixed methods study. *The Journal of continuing education in the health professions*. 2010; 30(4):221–8. [PubMed: 21171027]
27. Hazelton L, Allen M, MacLeod T, LeBlanc C, Boudreau M. Assessing Clinical Faculty Understanding of Statistical Terms Used to Measure Treatment Effects and Their Application to Teaching. *The Journal of continuing education in the health professions*. 2016; 36(4):278–83. [PubMed: 28350309]
28. Hernán MA, Hernandez-Diaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clinical Trials*. 2012; 9(1):48–55. [PubMed: 21948059]
29. Hernán MA, Robins JM. Per-Protocol Analyses of Pragmatic Trials. *New England Journal of Medicine*. 2017; 377(14):1391–8. [PubMed: 28976864]
30. Murray EJ, Hernan MA. Adherence adjustment in the Coronary Drug Project: A call for better per-protocol effect estimates in randomized trials. *Clin Trials*. 2016; 13(4):372–8. [PubMed: 26951361]
31. Toh S, Hernandez-Diaz S, Logan R, Robins JM, Hernán MA. Estimating absolute risks in the presence of nonadherence: an application to a follow-up study with baseline randomization. *Epidemiology*. 2010; 21(4):528–39. [PubMed: 20526200]
32. Hernán MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004; 15(5):615–25. [PubMed: 15308962]
33. Sullivan TR, Yelland LN, Lee KJ, Ryan P, Salter AB. Treatment of missing data in follow-up studies of randomised controlled trials: A systematic review of the literature. *Clin Trials*. 2017 1740774517703319.

What is new?

Key findings

- Patients prefer new-to-market medications only when substantially more effective or safer than existing choices.
- Patients prefer absolute risks in subgroups, and, when they expect to adhere, adherence-adjusted results, such as per-protocol effects.
- Investigators prefer both intention-to-treat and per-protocol effects but want better methods for per-protocol effect estimation.
- Pragmatic trials which estimated per-protocol effects used approaches which may result in bias, and no trials used adequate methods to adjust for loss to follow-up.

What this adds to what was known

- No clear guidance is available on the preferred types of causal information for medical decision-making.
- We assess preferences of patients and investigators towards causal effects of interest in pragmatic randomized trials.

What should change now?

- Pragmatic trials should focus on a goal of superiority in effectiveness or safety, rather than non-inferiority.
- Pragmatic trials should involve patients and patient advocates in specifying *a priori* subgroups to ensure relevance for shared decision-making.
- Absolute measures are the most interpretable and should be included in all trial reports.
- Per-protocol effects are of interest but clearer guidance on their estimation is needed, including appropriate adjustment for loss to follow-up.

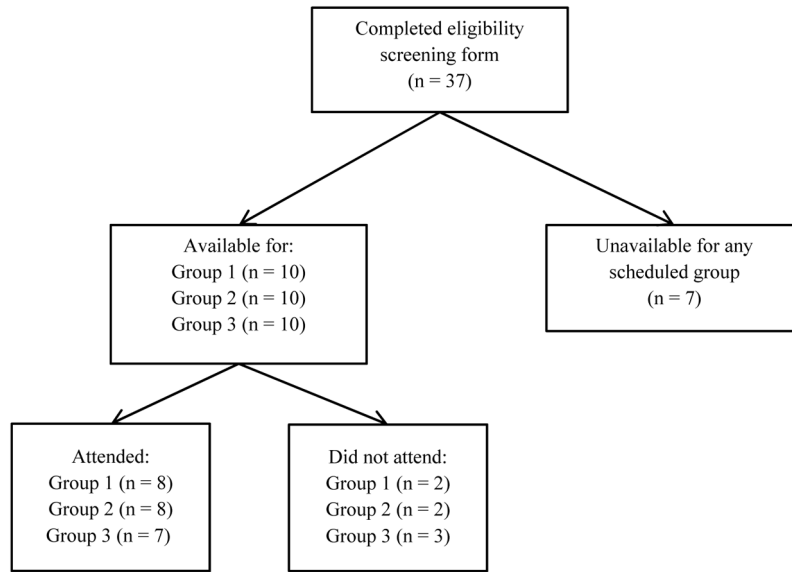


Figure 1.
Patient recruitment diagram

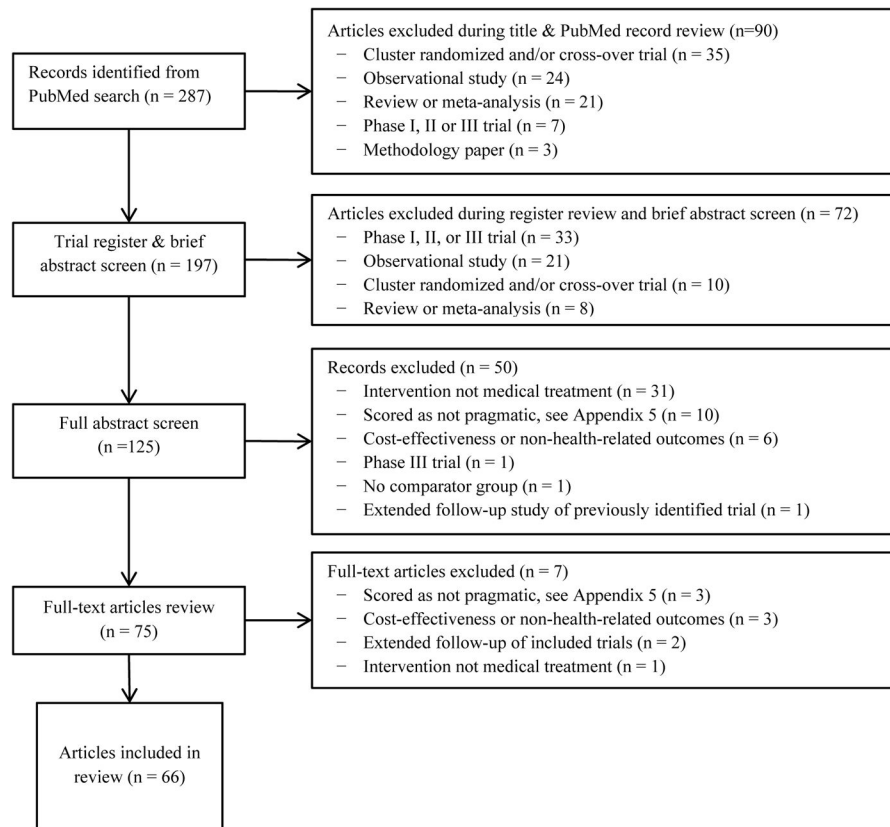


Figure 2.
Study selection diagram

Table 1

Characteristics of focus group, interview, and survey participants

Characteristics	Percent or Mean (SD)
Focus group participants (N=23)	
Gender	
Female	74%
Male	26%
Race	
African American	39%
White/Caucasian	61%
Ethnicity	
Hispanic	17%
Non-Hispanic	83%
Age	
18–29	4%
30–49	22%
50–64	44%
65 or older	26%
Highest level of education	
Some high-school	9%
High-school/GED	17%
Some college	9%
Bachelor's or Associate degree	39%
Master's or Doctoral degree	22%
Interviewed investigators (N=5)	
Gender (% women)	40%
Race (% white)	80%
Years of experience with clinical trials, mean (std)	16.6 (10.8)
Number of pragmatic trials conducted, mean (std)	3.2 (1.2)
Experience with other types of trials (%)	80%
Sample size of most recent trial, mean (range)	55,240 (300 to 180,000)
Duration of follow-up of most recent trial, years mean (range)	5.3 (1 to 20)
Surveyed investigators (N = 12)	
Years of experience with clinical trials, mean (std)	9.1 (9.1)
Number of pragmatic trials conducted, mean (std)	2.9 (3.1)
Experience with other types of trials (%)	83%
Sample size of most recent trial, mean (range)	14,468 (900 to 90,000)
Duration of follow-up of most recent trial, years mean (range)	1.6 (1 to 3)

Table 2

Characteristics of included pragmatic trials

Characteristic	Number of trials (n=63)
Study population	
Number of participants: median (IQR)	511 (241, 822)
Length of follow-up in months: median (IQR)	12 (6, 24)
% Female: median (IQR)	55% (40, 64)
% Female excluding women's health studies	49% (40, 60)
Race/ethnicity information reported	19 (30%)
Number of treatment arms	
2	39 (62%)
3	14 (22%)
4 or more	10 (16%)
Intervention	
Medication	18 (29%)*
Surgical protocol	13 (21%)
Medical device	12 (19%)*
Treatment protocol	8 (13%)
Counseling or therapy	5 (8%)
Diagnostic test	4 (6%)
Other	4 (6%)
Specialty	
Cardiovascular disease	8 (13%)
Infectious disease	7 (11%)
Reproductive health	7 (11%)
Primary care	7 (11%)
Orthopedics	6 (10%)
Psychiatry	5 (8%)
Other	23 (37%)
Primary outcome	
Symptom severity or resolution	38 (60%)
Mortality or survival (including live birth)	8 (13%)
Adverse events	4 (6%)
Health care utilization	4 (6%)
Biomarker	4 (6%)
Prevention	3 (5%)
Adherence	2 (3%)
Primary outcome type	
Continuous	28 (44%)
Binary	27 (43%)
Time to event	8 (13%)
Trial objective	

Characteristic	Number of trials (n=63)
Superiority only	45 (71%)
+safety	9 (14%)
+non-inferiority	2 (3%)
Non-inferiority only	6 (10%)
+safety	1 (2%)

* One trial assessed both medication and medical device

IQR: Interquartile range

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Representative quotations from focus groups and interviews supporting major themes

Major Theme	Source	Representative Quotation
1. Characterizing pragmatic trials	Investigator	<i>"Broadly, a pragmatic trial is one that attempts to address effectiveness in real world settings"</i>
2. Superiority vs non-inferiority	Patient	<i>"To me there has to be a point that they developed this drug. Like what else is going on with the drug? It's just as it is here, black and white. It's really not effective."</i>
	Investigator	<i>"A non-inferiority trial doesn't apply to this context because it is well-documented that usual care is very poor for this population. So a non-inferiority trial would be a waste of time...."</i>
3. Risk-benefit profile	Patient	<i>"I opted to take the standard medication because [the new medication] didn't say it had a 20-year research on it, and from what I've known in the past and heard, they really can't tell if it's better or not [unless] they've had some statistics, at least 20 years."</i>
	Investigator	<i>"The FDA doesn't demand causal analysis [per se] but wants to know the per-protocol [for adverse events] ... But the FDA is not so interested in the intention-to-treat, and they are willing to give up a little randomization to get the answer."</i>
4. Intention-to-treat vs per-protocol effects	Patient	<i>"It would depend on how critical the case was. If I had serious COPD and [...] both parents had died of it, I would say, 'You know what? I am committed to my health. I'm committed to taking it as prescribed.' So I'd be willing to try the new [less convenient] drug."</i>
	Patient	<i>"I would want to see the different groups, ethnic, race, sexes, weight, age, the whole spiel. I'd want to see all that first."</i>
	Investigator	<i>"So many selection factors determine adherence, so we need to adjust for them. We want to know the effect of intervention, not of being invited, so we need per-protocol effects."</i>
5. Characterizing effects: Absolute vs. relative risks	Patient (1 per 1000 vs 3 per 1000)*	<i>"No, I'd stay with the standard. The ratios are not much different ... between one out of a thousand and three out of a thousand, but the propensity for liver damage seems ... to exist and because the new drug hasn't had a lot of history, I'd be suspect of it."</i>
	Patient (1 per million vs 3 per million)	<i>"The liver damage ... would no longer bother me. To me, those are the same. I might go with the new drug if I really thought it was more effective and I needed it and I was trying everything else I could ... I wouldn't be concerned about this statistic at all."</i>
	Patient (1 per million vs 3 per million)	<i>"Because the rates, three out of one million, even though it's a million, it's still three out of a million and that one is always that chance, that risk, to me it's still a risk, being one out of the million. So I'm going to stick with that one standard."</i>
	Investigator	<i>"For advocacy, as a tool, sometimes relative measures are better. If we found 60% improvement [for example], for advocacy we might want to say that instead of 3% versus 4%... But as a scientist, absolute measures are the most honest."</i>
	Investigator	<i>"Particularly when dealing with dangerous results, we want to know if the absolute good effect is better than the absolute harm, so absolute measures are better [when assessing safety]."</i>

* Patients given choice of standard or new medication with risk of liver damage as specified

Table 4

Reasons for causal effect preferences and knowledge of analytic methods by effect estimates in their most recent pragmatic randomized trial among surveyed investigators

	Planned effect estimates in current or most recent pragmatic trial		
	Intention-to-treat as primary analysis (n=12)	Intention-to-treat as secondary analysis (n=3)	Per-protocol as secondary analysis (n=7)
Prefer this effect in an effectiveness (superiority) trial because:			
The causal question it answers	9	3	4
Always estimate this effect	6	1	0
Ability to estimate without bias	5	2	0
Ease of communicating to others	2	1	1
Ease of estimation	2	1	0
Trial-specific factors	2	0	0
Other	--	--	2
Familiarity with using post-randomization factors to adjust for loss to follow-up			
Not familiar	3	1	1
Somewhat familiar	8	2	5
Very familiar	1	0	1
Familiarity with using post-randomization factors to adjust for adherence			
Not familiar	7	3	3
Somewhat familiar	3	0	3
Very familiar	1	0	1

Table 5

Analytic designs of included pragmatic trials

Characteristic	Number of trials (n = 63)
Measures of occurrence for primary outcome	
Absolute risk, rate, or mean reported numerically for all trial arms	61 (97%)
Absolute risk, rate, or mean reported graphically only	2 (3%)
Measures of association for primary outcome	
Absolute difference measure only	25 (40%)
Relative comparison measure only	17 (27%)
Both absolute and relative measures	18 (29%)
Neither absolute nor relative measures: p-values only	4 (6%)
Primary effect of interest	
Intention to treat only	41 (65%)
+ Per protocol	22 (35%)
Statistical approach to estimating per protocol effect	
	(n=22)
Per-protocol population only [†]	20 (91%)
+ Statistical adjustment for non-adherence	1 (4%)
+ Other	1 (4%)
Sensitivity analyses for estimating per protocol effect	
	(n=22)
No, single approach used	17 (77%)
Yes, multiple approaches used (including multiple methods of defining per-protocol population)	5 (23%)
Statistical approach to loss to follow-up, for trials not using time to event outcomes	
	(n = 55)
Complete case only	31 (86%)
Multiple imputation	7 (13%)
Worst case imputation	3 (5%)
Complete case with sensitivity analysis: multiple imputation	4 (7%)
Complete case with sensitivity analysis: worst-case and/or best-case imputation	2 (4%)
Last observation carried forward or linear interpolation	3 (5%)
Imputation, method not specified	1 (2%)
Best case imputation	1 (2%)
Unclear or not reported	3 (5%)

[†]Per-protocol population analysis is conducted by restricting to those individuals who adhered to their trial assignment with no additional statistical adjustment for selection into this subsample.

Table 6

Characterizing patient preferences for effect measures

Scenario	% of Patients:			
	Preferring standard medication	Preferring new medication	Preferring neither	Requesting more information
“The new medication is 3 times more likely to cause liver damage than the standard medication”	100%	0%	0%	0%
“1 out of a million people who take the standard medication get liver damage; and 3 out of a million people who take the new medication get liver damage”	57%	30%	0%	14%
“1 out of 1000 people who take the standard medication get liver damage; and 3 out of 1000 people who take the new medication get liver damage”	96%	0%	0%	4%
“1 out of 10 people who take the standard medication get liver damage; and 3 out of 10 people who take the new medication get liver damage”	82%	0%	18%	0%