NIH Collaboratory Rethinking Clinical Trials®
Health Care Systems Research Collaboratory

Promoting Both Internal and External Validity: Designing the Trial to Match Its Intention

Contributors

Merrick Zwarenstein, MBBCh, MSc, PhD
Professor, Department of Family Medicine
Epidemiology & Biostatistics
Western University
Senior Core Scientist
IC/ES Western Primary Care & Health Systems Research Program
London, Ontario, Canada

Ahmed Al-Jaishi, PhD
Health Research Methodology
McMaster University
Hamilton, Ontario, Canada

Amit X. Garg, MD, PhD
Professor, Department of Medicine
Epidemiology & Biostatistics
Western University
Senior Core Scientist, Site Director, and Program Lead
IC/ES Western Kidney Dialysis & Transplantation Research Program
Nephrologist, London Health Sciences Centre
Victoria Hospital
London, Ontario, Canada

The greatest strength of randomized controlled trials (RCTs) is that *randomization* distributes known and unknown confounders equally between arms of the trial and increases the findings' internal validity. Internal validity means that the point estimate of the effect size of the intervention in comparison with the comparator(s) from that RCT is unbiased, but only for the actual participants in the RCT itself. No matter how well an RCT implements strategies to increase internal validity, these strategies do not help with the external validity of the findings of that RCT. External validity is the applicability of the findings of an RCT to other potential RCT participants present at the time and place of the

trial but who did not consent, meet trial inclusion criteria, were lost to follow up, were those at that same treatment center but a later time, or were those whose care is provided elsewhere in a different setting or context.

The internal and external validity of RCT findings are independent of each other: internal validity arises entirely from randomization and its proper implementation (allocation concealment) and maintenance (minimal loss to follow up), while external validity is always a judgment call based on the degree of similarity between the participants and contexts of the RCT and those in the situation to which the RCT findings will be applied (by clinicians, patients, policymakers, or other relevant real-world decision-makers).

Although internal validity on its own does not ensure the usefulness of an RCT finding for decision-makers in other contexts, it is still absolutely needed as a foundation for using the findings of any trial. For example, an RCT conducted in a similar context to some real-world situation and with comparable participants is useless for decision-makers if randomization was inadequate and undermined the internal validity of that RCT. What, then, is the usefulness of traditional RCTs that focus on internal but not external validity? Even if not directly useful for decision-makers due to uncertainty about external validity, an internally valid trial helps test a hypothesis that an intervention causes some particular change in some outcome. In other words, a test for the existence (or not) of a mechanism of action. This laboratory-like randomized causal experiment, between groups formed to be comparable by randomization that receives the intervention under evaluation or not, with variability minimized by selection of a homogeneous set of participants and rigidly constrained procedures permit confident assertions that the causal mechanism exists (or not) and thus contributes to scientific understanding. Although we accept this as *generalizable* scientific knowledge (i.e., "the causal mechanism exists in study participants with this problem, and probably exists also in others of this species"), we do not expect the effect size in such an experiment to be directly applicable to others with this problem, perhaps with different severity, or additional problems, in less laboratory-like, real-world contexts.

This set of basic assumptions about internal validity and generalizability, external validity and direct applicability, is commonplace among statisticians and is less well understood by clinicians. In their influential publications (book and paper), Schwartz and Lellouch skipped over any explanation of this basic understanding and started at the next step: that there are two distinct purposes that a randomized trial can serve: (1) generalizations regarding the existence of a causal mechanism, which they term *explanatory*, or (2) specific findings on effect size within a real-world context that can, with caution, be directly applied to a decision in another context, which they term *pragmatic* (Schwartz and Lellouch 1967). They then assert that these two purposes (sometimes called intentions, approaches, or attitudes) require different design choices.

The purpose should be decided before embarking on designing a trial, and each element of the trial design should be aligned to the chosen purpose. If the purpose is explanatory, then the design choices would minimize any factor other than the intervention itself that might cause the outcome to vary; this gives the causal mechanism its maximum room for effect. If the purpose is pragmatic, then the design choices would favor participants, clinicians, settings, and other conditions as close as possible to the unconstrained usual care situation expected for that intervention in the future. This approach provides decision-makers with a direct estimate of effect size to adjust for any residual differences between their situation and the trial.

A pragmatic trial would retain the advantage for the real world even if the trial were conducted in a usual care context different from that of the decision-makers. This is because more explanatory trials avoid the constraints in the delivery of care, selection and recruitment of participants, clinicians, comparators, setting, outcome, analysis, and follow-up used to create the "laboratory-like conditions" that minimize sources of variability in outcome. These constraints make it easier to detect causal differences arising from the intervention alone. However, they may also distort the real-world context so much that decision-makers reading the trial results will struggle to apply the findings and choose between alternative interventions in their usual care context.

Traditional RCTs have tended toward the explanatory approach, resulting in a design that specifies (and depends on) these characteristics:

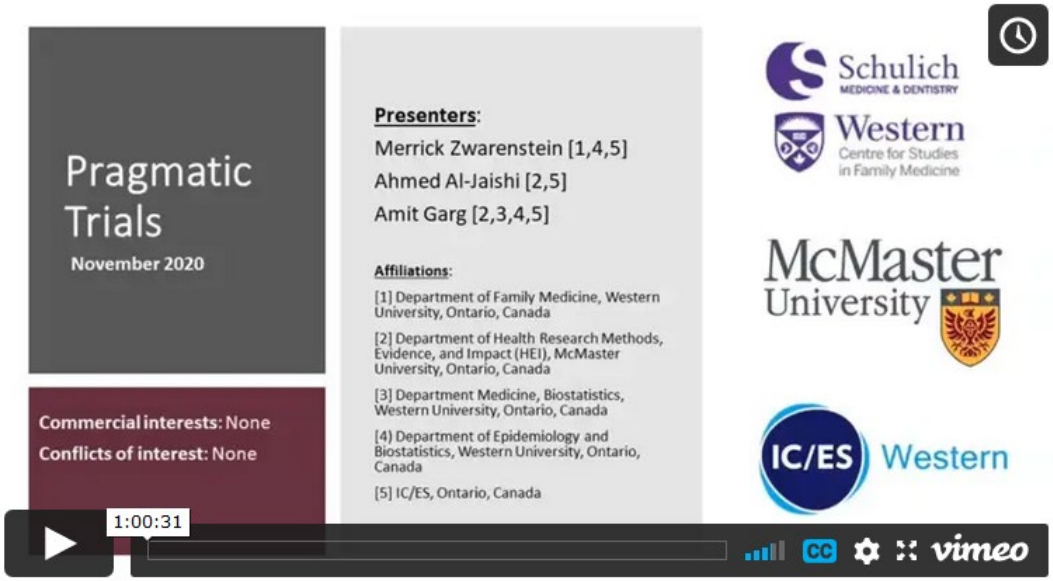
- Excluding all but a narrow and homogeneous set of patients, care providers, and settings
- Optimizing care delivery so that it is more consistent than could be expected for usual care for the interventions under evaluation
- Collecting outcomes that document mechanisms and underlying processes rather than events that are important to patients ("patient-centered")
- Comparing with control groups receiving difficult-to-interpret comparators such as placebo

While these choices help answer a question on the mechanism of action, they are less likely to produce evidence that directly supports a decision to use the newly tested intervention. Precisely because such trials are conducted under controlled and specialized conditions, their external validity for making decisions may be compromised (Kennedy-Martin et al., 2015).

Focusing on the trial's intention is the first step in designing a trial that successfully answers its intended primary research question. While there is a contrast between a pragmatic intention and an explanatory intention, there is no dichotomy in the design choices made to match each intention. Instead, trials will vary across the spectrum of design decisions leaning toward choices that match the trial's purpose. Differences in these intentions require specific design choices, and the PRECIS-2 tool can help investigators design their trial to align with its intention. These points are illustrated in the *Living Textbook* section PRECIS-2 Case Study.

Additional Resources

Hear the coauthors present on PCT Grand Rounds, November 13, 2020: Pragmatic and Explanatory Attitudes to RCTs: Using the PRECIS-2 Tool to Describe the Design of the MyTEMP Trial



References

Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. 2015. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials. 16:495. doi:10.1186/s13063-015-1023-4. PMID: 26530985.

Schwartz D, Lellouch J. 1967. Explanatory and pragmatic attitudes in therapeutical trials. J Chronic Dis. 20:637-648. doi:10.1016/0021-9681(67)90041-0. PMID: 4860352.