

## Five good reasons to be disappointed with randomized trials

### Background

Randomized controlled trials (RCTs) are recognized to exhibit very high levels of evidence, representing a coveted position near the top of the evidence-based pyramid [1]. Both authors of this editorial have been part of small to large-scale RCTs and support the need for this form of research design. Yet, few things annoy us more than the *deification* that clinicians and selected researchers have given to randomize controlled trials. Yes, RCTs are useful in testing the efficacy and effectiveness of interventions between groups; essentially, identifying which treatment intervention is superior *between* two or more unique groups [2]. Moreover, RCTs are necessary to reduce bias and confounding and are perceived to yield causal inferences [3]. *However* (and we can't emphasize this enough), it is our impression that few understand the noteworthy limitations of RCTs, and even fewer are able to extrapolate how these limitations influence clinical practice. Our experiences with these misunderstandings have prompted us to outline *some* (trust us, there are more) of the limitations of RCTs, specifically those that might influence clinical practice in an orthopedic setting.

### Limitations

*Reason One: Right Question-Wrong Design:* A common response we hear is the belittling of a given study finding because it didn't involve an RCT. It is imperative to understand that RCTs are a form of research design and this design is not appropriate for all forms of research needs. For example, diagnostic accuracy studies are best analyzed using a case-based, case-control design. Rare diseases are best studied using case-control designs. If one is looking at predictive analytics then a prospective cohort design is the design of choice [2]. Looking for patterns and effects across different data sources?; a systematic review or a meta-analysis is the design of choice. And although an influential paper from 2004 called for better reporting of harms in RCTs [4,5], an RCT is not the most appropriate study design to truly understand the prevalence of these adverse events [6]. An observational case-cohort design will better reflect the population, prevalence and downstream influence of harms associated with dedicated care processes [7].

*Reason Two: The Marginal Patient:* Perhaps the most well-known limitation of an RCT is external validity. External validity is the degree to which the conclusions in your study would hold for other persons in other places and at other times. In RCTs, there are unavoidable disparities between the study conditions and populations in comparison to the conditions and populations in which the finding will be inferred [8]. A common assumption is that the findings would be transferable to all patient populations, treatment environments, and cultures. This 'it-works-somewhere' [9] concept is defined as: *projected realism*.

In an effort to 'control' for confounding variables and increase study power, a homogenous sample of diagnostically uniform patients are included that may not represent the actual demographics and complexity in the clinic. These less simple patients are termed 'the marginal patients' because the average patient may or may not respond to a given treatment [10–12]. Unfortunately, many of the requirements needed in an RCT to improve internal validity (and control for confounding bias) result in an artificial-like setting that does not closely match a real-world environment [13]. Despite the notable juxtaposition between external and internal validity, many RCTs and observational designs involving similar interventions and participants find similar results [14]. Because RCTs are often exceptionally expensive, authors have recommended different designs, alternative data sources, and unique methodological approaches to identify similar findings (at a reduced cost) [15].

*Reason Three: Mixed Treatment Effect-* Just because one group reports better outcomes than another group in an RCT, it does not mean that the intervention in the group with better outcomes works for all individuals in that group or future groups [13]. Yes, if one finds differences between two groups, the intervention that is associated with an improved outcome may indeed have higher efficacy (for the group tested). Nevertheless, as most studies demonstrate, some individuals in both groups improve whereas some individuals in both groups do not. An RCT only functions to show whether more people improved in one group versus the other, or 'who' (which group) benefits. *Why* someone improved is not a property of a RCT.

To determine 'why' someone improves requires a causal mediation design. Causal mediation analysis identifies potential pathways that could explain *why*

the outcomes were more effective with that intervention [16]. Causal mediation analysis allows an understanding of the roles of intermediate variables that lie in the causal path between the treatment and outcome variables, and allows the clinician to focus on both the mediating and primary (intervention) variables with targeted applications. Additionally, not all patients may be appropriate to a given mix of interventions with similar conditions. Thus, determining an effective treatment mix may provide more clinically useful information as opposed to a single treatment approach that demonstrates an effective average treatment effect [17–19]. Sadly, although causal mediation designs are often secondary analyses within an RCT, an RCT in isolation does not provide that information.

*Reason Four: Treatment Fidelity:* Intervention fidelity refers to the reliability and validity of the clinical interventions that are used in the randomized trial [20]. In other words, fidelity reflects the applicability of the interventions for the condition of interest, whether the interventions are appropriately performed (application, dosage, and intensity) and whether the interventions adequately represent how the intervention is performed in clinical practice. Interestingly, past studies have found that intervention fidelity is consistently either poorly performed, poorly reported or both [21]. Unfortunately, because of the costs associated with RCTs, fidelity is commonly sacrificed. Even pragmatic randomized trials (trials designed to test the effectiveness of the intervention in a broad routine clinical practice) are guilty of limited fidelity in the application of behavioral or exercise-based interventions [20].

*Reason Five: Unmeasured Bias:* The post-randomization experience is the period that immediately follows individuals' consent and randomization to one of the treatment groups [22]. Randomization is used to reduce errors, differences in groups, and confounding properties that are unforeseen. The post-randomization experience ('what happens after the randomization') can also be a period in which bias may play a notable role. Outside of fidelity and some of the aforementioned items, there are *five major considerations involving the post-randomization experience*. The *Hawthorn* effect is a change in behavior of the research subjects, administrators, and clinicians in experimental or observational studies [23]. Patients hold certain beliefs and expectations regarding a treatment that have been shown to influence the outcomes [24]. If the allocated treatment group does not match the *patients' beliefs and expectations* then the treatment effect is likely subdued. *Personal equipoise* exists when a clinician has no good basis for a choice between two or more care options or when one is truly uncertain about the overall benefit or harm offered by the treatment to his/her patient [25]. *Mode of administration bias* exists when the method of outcomes collection (how outcomes were collected from the

research participant) is tainted between clinician and research subject [26]. Lastly, *contamination bias* occurs when the members of one group in a trial receive the treatment or are exposed to the intervention that is provided to the other group.

To reinforce the influence of the Hawthorne effect and personal equipoise, we provide the following *examples*. First, provider, health services patterns, and comparison of profession are study foci that are particularly predisposed to the Hawthorn effect. Although the studies involve randomizing to control biases, *clinician behaviors* are likely to change since they know they are being evaluated in a formal study. For example, if you are the prescribing physician in a trial that is examining the negative effects of opioids, you are likely going to prescribe fewer opioids. Personal equipoise toward a particular intervention will unconsciously cause an improve outcome for the treatment of preference. For example, in randomized trials where clinicians preferred a particular treatment approach (despite being randomized between two groups), the preference influenced outcomes in a way that supported their preference [27,28].

## Summary

Randomized controlled trials are useful in testing the efficacy and effectiveness of interventions between groups [2]. Understanding their limitations is essential before extrapolation to clinical practice. Other research designs are needed to understand the diagnosis, validity of outcomes, and other important research issues. Participants enrolled in RCTs may or may not adequately represent the full population in which the study is designed to represent. Randomized controlled trials evaluate the effects of treatment at population levels and do not explain *why* the outcomes were more effective with that intervention [9]. The care provided may or may not reflect what is appropriately provided in clinical practice. And lastly, a biased post-randomization experience is not protected by the initial randomization. Careful controls are necessary at this phase of the trial as well.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References

- [1] Murad MH, Asi N, Alsawas M, et al. New evidence pyramid. *BMJ Evidence Based Med.* 2016;21(4).
- [2] Fritz JM, Cleland J. Effectiveness versus efficacy: more than a debate over language. *J Orthop Sports Phys Ther.* 2003;33:163–165.
- [3] Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med.* 2018;210:2–21.

- [4] CONSORT Group, Ioannidis JP, Evans SJ, Gøtzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med.* 2004;141(10):781–788.
- [5] Chan AW, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med.* 2013;158(3):200–207.
- [6] Zorzela L, Golder S, Liu Y, et al. Quality of reporting in systematic reviews of adverse events: systematic review. *BMJ.* 2014;348:f7668.
- [7] Checkoway H, Pearce N, Kriebel D. Selecting appropriate study designs to address specific research questions in occupational epidemiology. *Occup Environ Med.* 2007;64(9):633–638.
- [8] Pearl J. Challenging the hegemony of randomized controlled trials: a commentary on Deaton and Cartwright. *Soc Sci Med.* 2018;210:60–62.
- [9] Mulder R, Singh AB, Hamilton A, et al. The limitations of using randomised controlled trials as a basis for developing treatment guidelines. *Evid Based Ment Health.* 2018;21(1):4–6.
- [10] McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial-infarction in the elderly reduce mortality - analysis using instrumental variables. *JAMA.* 1994;272(11):859–866.
- [11] Brooks J, McClellan MM, Wong HS. The marginal benefits of invasive treatments for acute myocardial infarction: does insurance coverage matter? *Inquiry.* 2000;37(1):75–90.
- [12] Harris KM, Remler DK. Who is the marginal patient? Understanding instrumental variables estimates of treatment effects. *Health Services Res.* 1998;33(5):1337–1360.
- [13] Gelman A, Loken E. The statistical crisis in science. *Am Scientist.* 2004;102:460–465.
- [14] Ioannidis JPA. Randomized controlled trials: often flawed, mostly useless, clearly indispensable: a commentary on Deaton and Cartwright. *Soc Sci Med.* 2018;210:53–56.
- [15] Frieden TR. Evidence for Health Decision Making – Beyond Randomized, Controlled Trials. *N Engl J Med.* 2017;377(5):465–475.
- [16] Rudolph KE, Goin DE, Paksarian D, et al. Causal mediation analysis with observational data: considerations and illustration examining mechanisms linking neighborhood poverty to adolescent substance use. *Am J Epidemiol.* 2018. [Epub ahead of print].
- [17] Bernstein J. Not the last word: choosing wisely. *Clin Orthop Relat Res.* 2015;473(10):3091–3097.
- [18] McCulloch PM, Nagendran WB, Campbell A, et al. Strategies to reduce variation in the use of surgery. *Lancet.* 2013;382(9898):1130–1139.
- [19] Birkmeyer JD, Reames BN, McCulloch P, et al. Understanding of regional variation in the use of surgery. *Lancet.* 2013;382(9898):1121–1129.
- [20] Cook CE, George SZ, Keefe F. Different interventions, same outcomes? Here are four good reasons. *Br J Sports Med.* 2018;52(15):951–952.
- [21] Toomey E, Currie-Murphy L, Matthews J, et al. Implementation fidelity of physiotherapist-delivered group education and exercise interventions to promote self-management in people with osteoarthritis and chronic low back pain: a rapid review part II. *Man Ther.* 2015;20:287–294.
- [22] Choudhry NK. Randomized, Controlled Trials in Health Insurance Systems. *N Engl J Med.* 2017;377(10):957–964.
- [23] Sedgwick P, Greenwood N. Understanding the Hawthorne effect. *BMJ.* 2015;351:h4672.
- [24] Harris J, Pedroza A, Jones GL. Predictors of pain and function in patients with symptomatic, atraumatic full-thickness rotator cuff tears: a time-zero analysis of a prospective patient cohort enrolled in a structured physical therapy program. *Am J Sports Med.* 2012;40(2):359–366.
- [25] Cook C, Sheets C. Clinical equipoise and personal equipoise: two necessary ingredients for reducing bias in manual therapy trials. *J Man Manip Ther.* 2011;19(1):55–57.
- [26] Cook C. Mode of administration bias. *J Man Manip Ther.* 2010;18(2):61–63.
- [27] Cook C, Learman K, Showalter C, et al. Early use of thrust manipulation versus non-thrust manipulation: a randomized clinical trial. *Man Ther.* 2013;18(3):191–198.
- [28] Bishop MD, Bialosky JE, Penza CW, et al. The influence of clinical equipoise and patient preferences on outcomes of conservative manual interventions for spinal pain: an experimental study. *J Pain Res.* 2017;10:965–972.

Chad E. Cook

*chad.cook@duke.edu*, Department of  
Orthopaedics, Duke University, Durham, NC, USA  
Duke MSK Group, Duke Clinical Research Institute,  
Durham, NC, USA

 [chad.cook@duke.edu](mailto:chad.cook@duke.edu)

Charles A. Thigpen

ATI Physical Therapy, Department of Clinical  
Excellence, Greenville, SC, USA  
Arnold School of Public Health, Center for  
Effectiveness in Orthopedic, University of South  
Carolina, Greenville, SC, USA