Beyond Randomized Clinical Trials: Emerging Innovations in Reasoning About Health

JODI SCHNEIDER
School of Information Sciences, University of Illinois at Urbana-Champaign, USA
jodi@illinois.edu

SALLY JACKSON
Department of Communication, University of Illinois at Urbana-Champaign, USA
sallyj@illinois.edu

Specialized fields may at any time invent new inference rules—that is, new warrants—to improve on their stock of resources for drawing and defending conclusions. One such invented warrant, Randomized Clinical Trial, is widely regarded as the gold standard for making inferences about causal relationships between medical treatments and patient outcomes. Tensions that arise from the competing perspectives of scientists, clinicians, and patients have recently led to reconsideration of RCT and to emergence of alternative research strategies, notably 'pragmatic trials' and 'N-of-1 trials'.

KEYWORDS: field-specific reasoning, medical reasoning, N-of-1 Clinical Trial, Pragmatic Clinical Trial, Randomized Clinical Trial, warrants, warrant-establishing arguments

1. INTRODUCTION

Toulmin (1958) pointed out the possibility that specialized fields may at any time invent new inference rules—that is, new warrants—to improve on their stock of resources for drawing and defending conclusions. This appears to be happening at a very rapid pace in the field of health science, where several waves of innovation have occurred over the past century or more. Jackson and Schneider (2018) analyzed one recent innovation, a form of evidence aggregation known as a Cochrane Review. Although Cochrane Review functions argumentatively as a generalized warrant, it has special features not normally attached to warrants, including technical components invented specifically to support the use of the warrant in reasoning within the field. We introduced the term "warranting devices" for a class of such innovations that involve an inference rule packaged with its technical components in such a way that any use of the rule includes tacit assurance that it generates dependable conclusions. A warranting device, then, is a specialized inference rule, invented within a field for some particular argumentative purpose, and backed by a set of assurances that may be partly material, partly procedural, and even partly institutional.

In subsequent work, Schneider and Jackson (2018b) examined another warranting device known as the Randomized Clinical Trial (RCT), widely regarded as the gold standard for making inferences about causal relationships between medical treatments and patient outcomes. Still controversial through the early twentieth century, RCT achieved broad acceptance within the field

as a result of warrant-establishing arguments circulating in the medical literature starting in the 1950s (Schneider & Jackson, 2018a). In this paper, we examine several less well-established movements within health science (notably 'pragmatic trials' and 'N-of-1 trials') that seek to go beyond RCT as a basis for reasoning about treatments. We consider how early decisions about the design of the warranting device (notably, a focus on group averages as central to inference about cause and effect) brought about undenied improvements in reasoning while also sowing seeds for later dissatisfaction with how results were translated into clinical practice.

Although any proposed warranting device may be established through successful demonstration that it can produce dependable conclusions, these devices are by their very nature changeable, either wholly or in part. A device may become stabilized within the reasoning practices of a field at one point in time, then de-stabilized at a later point in time, because new vulnerabilities in the device are discovered, because some new device pushes an older one toward obsolescence, or because the arguments generated by the device meet new forms of criticism in new discourse contexts. Warrant-establishing argument is never completely conclusive; disagreement over the acceptability of an invented warrant can always be re-opened.

In this paper, we explore the arguments that have helped to re-open debate over RCT, exploring the tensions that arise from the competing perspectives of scientists, clinicians, and patients.

2. CLINICAL TRIALS

Clinical trials have become, or are quickly becoming, a worldwide standard for generating evidence of the effects of proposed treatments. The feasibility of clinical trials depends on material and institutional resources. For instance, they are affected by the health care systems in a locality (e.g., the logistics of recruiting patients and managing a controlled administration of treatments). They may also be subject to different restrictions in different national jurisdictions. Both the feasibility of clinical trials and the quality of evidence resulting from them can be affected by societal conditions that are outside the control of scientists, requiring well-organized efforts to create conditions more supportive of clinical trials (e.g., for the EU, making cross-national recruitment feasible; see Demotes-Mainard, & Kubiak, 2011). Clinical trials have economic value, and globalization of the practice is thus partly driven by the pharmaceutical industry (Thiers, Sinskey, & Berndt, 2008).

While the basic logic of clinical trials is global in reach, national or regional institutional context is important to understanding how clinical trialing as a practice has developed. In the US, clinical trials often depend upon hospitals that have a research mission, especially university hospitals. Funds needed for independent (non-industry) research are controlled by the US National Institutes of Health. The US Federal Drug Administration governs approval of new pharmaceuticals and has played a significant role in the institutionalization of the phase structure of clinical trialing (described below). Finally, journal editors can exert pressure on how research is conducted by limiting publication opportunities based on compliance with both scientific and ethical standards (as in De Angelis et al., 2004, and Taichman et al., 2017). These institutional actors, pushing toward their own goals, and sometimes pushing back against one another, have shaped the structure of clinical trials in the US.

Clinical trials are defined by the NIH as research studies "in which one or more human subjects are prospectively assigned to one or more interventions (which may include placebo or other control) to evaluate the effects of those interventions on health-related biomedical or

behavioral outcomes" (U.S. National Institutes of Health, 2014). They are experiments on human subjects whose independent variables are potential treatments and whose dependent variables are aspects of health or well-being. The logic of clinical trials is apparent from the diagram in Figure 1, showing random assignment of a large number of patients to contrasting forms of treatment. Inferences about whether and how the treatments differ in effects are delegated to tests of statistical significance and quantitative measures of effect size.

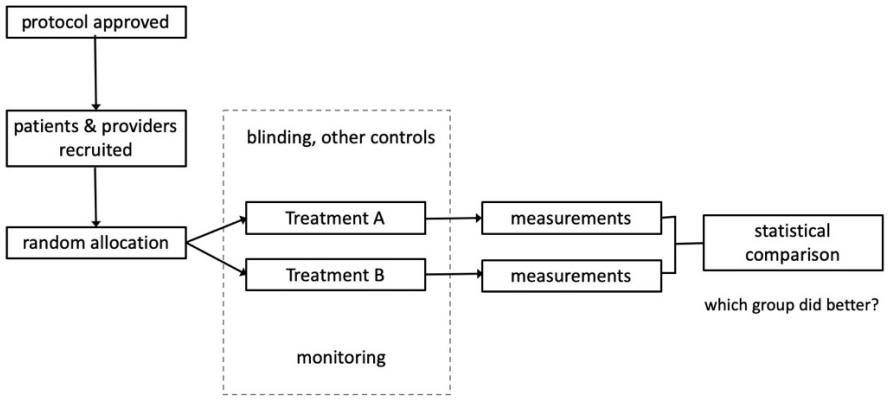

Figure 1 – A simple Randomized Clinical Trial (Schneider & Jackson, 2018a).

Conducting experimental research on human subjects in *phases* allows for (and often requires) evaluation of the safety of a treatment for healthy patients prior to evaluation of the efficacy of the treatment for sick patients. In the highly regulated world of pharmaceuticals, conducting research in phases has become institutionalized to such an extent as to permit explicit codification by the US Federal Drug Administration (Office of the Commissioner, 2019). Clinical research (on human subjects) is expected to begin with a demonstration that the drug can be safely given to humans. For new drugs, they must first have been tested on non-human animals (Center for Drug Evaluation and Research, 2019). Phase I trials recruit healthy subjects, typically not very many, and may involve such design features as dose escalation over the course of the trial. Assuming that a safe dose level is found, the drug may be used in a Phase II trial on volunteers from the relevant patient population, allowing not only continued assessment of safety for this patient population but also assessment of efficacy. Phase III trials are larger in size (number of patients) and longer in duration, to allow for 'small' effects to be detected, especially any adverse effects that may not be noticed in a smaller sample or over a shorter period of time. A drug that assembles evidence of safety and efficacy over these three phases is a good candidate for FDA approval. From an argumentative perspective, FDA approval encapsulates claims of safety and *potential* benefit for patients to whom the drug may be administered. Post-approval clinical trials are known in the FDA world as Phase IV trials and have aims similar to Phase III.

The distinction between Phase I and the other phases is particularly significant for drug treatments and certain other interventions: Phase I trials recruit healthy patients, not those with the condition that the intervention is expected to treat, so they do not normally provide much reason to believe that an intervention will be beneficial for treating that condition. From an argumentative perspective, promising results from a Phase I trials do not even provide evidence that the treatment

is *safe* for patients with that condition. They do, however, provide evidence on safety that can allow prospective volunteers for Phase II trials to consider their own risk realistically. Phased trialing adds considerable nuance to what claims are supported by RCTs; an RCT may establish a safe dose level, or a difference between one treatment and another, or a certain "success" rate in patient care, and although all of these are labelled 'effects' of treatments in their respective studies, they are not all the same. It takes a very long time to get through all of the work of a phased trial sequence, and at any point in time, the kind of claims that are actually warranted vary by which phase has or has not been passed.

Understanding how thoroughly intertwined clinical trials are with institutional context is very important to understanding how they can warrant inferences about treatment effects—and especially important for understanding why publics push back against them when these inferences become part of reasoning about actual treatment of actual patients.

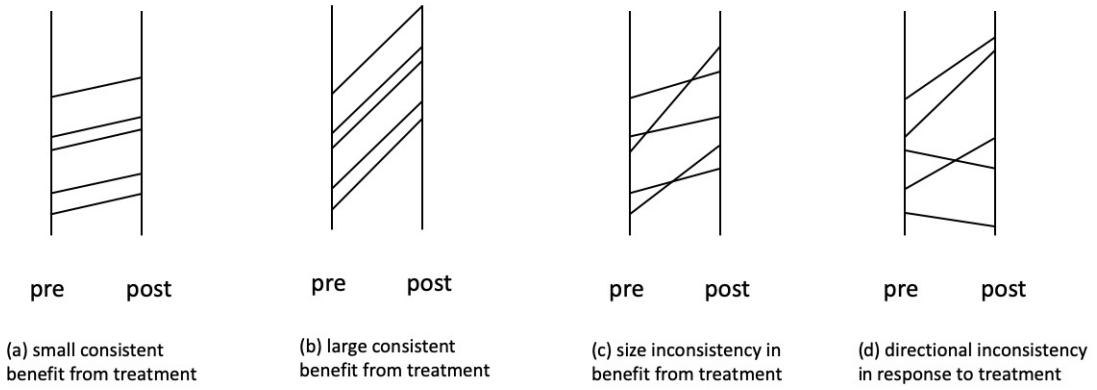

Figure 2 – Treatment effect at individual level expressed as change from pre-test to post-test, with each line representing one patient. Lines sloped upward represent benefit from treatment; lines sloped downwards represent worsened condition after treatment.

Clinical trials answer very well-defined questions that are relevant and important to decisions about how to treat patients, but practitioners and researchers alike know that successful results in Phases I to III do not assure that the drug or other treatment evaluated will be successful for all patients under all conditions. In fact, many of these sequences produce evidence that treatment effects vary widely from patient to patient. An average benefit may appear when some patients benefit while others do not, or when some benefit while others are actually harmed. Statistically, this situation is known as 'person by treatment interaction,' but it is not always visible (or estimable) within a standard clinical trial. To estimate person by treatment interaction, a researcher must observe what happens to each individual, both with and without the proposed treatment (for example, in a pretest/posttest design). In Figure 2, each panel shows the effect of treatment as change from a pretest measurement (without the treatment) to a posttest measurement (with the treatment). Each patient's pretest and posttest measurements are represented by a line drawn between two vertical axes. Lines with positive slope (rising to the right) represent patients who benefitted; lines with negative slope (falling to the right) represent patients who did worse after being treated. Variability in the slopes of the lines represent person by treatment interaction. All of these cases produce an average benefit; all four configurations could produce a statistically

significant benefit for the treatment (especially if only a small proportion of patients do worse with the treatment than without). For patients and their care providers, this means that a treatment that is beneficial on average may or may not be beneficial for any one individual. Likewise, when one treatment is shown to be better than another on average, it may still be true that the "less effective" treatment is best for some patients.

So even after Phase III, there can remain a large gap between what is established through this trial sequence and what a reasonable physician or patient would want to know before choosing to administer or to accept the treatment, either as a standard option or—especially—as a specific choice for an individual patient. As rules and preferences are imposed over time by funders, regulators, and publishing gatekeepers, this gap can widen—or narrow.

But the gap has become more noticeable over time. An under-appreciated fact is that the "logic" behind a particular innovation in inference, even when made quite explicit, cannot always be fully evaluated without applying it to the task of drawing conclusions. After the initial successful defense of RCT for drawing conclusions about medical treatments, there was great optimism about its potential and great momentum behind exploiting this potential. But as medical practice has become more infused with evidence from RCTs, what seemed like unproblematic reasoning has turned out to have unexpected limitations. Hundreds of Cochrane Reviews framed by practical questions about care locate *zero* papers reporting data worth aggregating.[1] Evidence worth aggregating based on each review's pre-specified criteria is not always forthcoming, either: some reviews remain empty for ten years or more, even after repeated attempts to locate relevant evidence (Yaffe, Montgomery, Hopewell & Shepard, 2012). This suggests gaps between the answers health care practitioners want and the evidence available for synthesis from RCTs and other methods. For example, what we know scientifically about possible treatments for a health condition is dependent in part on what it is allowable to study, in part on what is prioritized by funding sources, and in part on what scientists themselves find interesting. The lack of scientific evidence for something is often a direct consequence of institutional actors having no interest in it.

3. PRAGMATIC TRIALS

RCT "technology" might have developed quite differently than it actually has—which is to say that its core ideas could have been elaborated in multiple different ways. Bradford Hill's defense of controlled clinical trials in the 1950s, analyzed by Schneider and Jackson (2018a), left many avenues of development open—not just the avenue that has resulted in NIH's three (or four if counting post-approval Phase IV) distinct trial phases.

Early proponents of alternative technological directions included Daniel Schwartz and Joseph Lellouch, whose 1967 article titled "Explanatory and Pragmatic Attitudes in Therapeutic Trials" took decades to attract a large enough following to get 'pragmatic trials' broadly acknowledged as a fourth phase. Schwartz and Lellouch argued that a basic inferential strategy of comparing outcomes obtained with contrasting treatments could be undertaken with either a purely epistemic aim as in basic science (to explain something) or with a pragmatic, choice-oriented aim (evaluating a course of

[1] Roughly 9% of Cochrane reviews are empty. This ratio seems relatively constant over time: As of August 15, 2010 Yaffe and colleagues (2012) found 8.7% empty reviews (376 of 4320 reviews), while we determined that as of January 3, 2018, 9.2% (659 of 7156) published Cochrane reviews in the Cochrane Library were empty.

treatment or choosing a treatment policy). Both aims can be served by a standard experimental design (shown earlier in Figure 1): People are randomly allocated to one of two alternative treatments, Treatment A or Treatment B; measurements are taken (and statistically compared) on whatever physical or mental state Treatments A and B are expected to improve.

Schwartz and Lellouch pointed out that despite commonality of this structure, designing a trial to satisfy explanatory aims is very different from designing a trial to satisfy pragmatic aims. We will not review all of the nuance of their argument but simply summarize three issues that clearly differentiate explanatory and pragmatic aims: how to form comparison groups, how to conceptualize treatments, and how to select meaningful outcomes.

3.1 Comparison groups

That comparison groups should be formed at random from a common pool is not disputed by Schwartz and Lellouch. Their concerns are with how the common pool is developed, and with what happens when individuals from this common pool drop out after random assignment to a treatment. They argue that in such cases, statistical analysis may be conducted either on the premise that the dropouts are simply people for whom the treatment was unsuitable (that is, people who have nothing to tell us about the potential efficacy of the treatment), or on the premise that the treatment is problematic in some way (by virtue of failing for some of those it aims to benefit). As they put it, "in the first [explanatory] case the class of patient is defined to fit the predetermined treatments, while in the second [pragmatic] the treatments are defined to fit the predetermined class of patients" (p. 643).

3.2 Treatments

When two proposed treatments are to be compared, it will normally be the case that each considered individually is a complex assembly of components, including the form in which the treatment would most conveniently be administered, the time over which it would typically be administered, the setting in which it would ideally be administered, and much more. The explanatory attitude strives toward a contrast in which as many of these components as possible are equalized between the treatments to be compared, while a pragmatic attitude strives for a contrast between the optimal arrangement for each of the treatments. Conducting the comparison between two (artificially) equalized treatments invites the possibility that neither treatment works up to its potential. Conducting the comparison between two optimized treatments allows for all manner of confusion over *exactly what* makes the better of the two treatments better.

Suppose, for example, two different substances have been approved for treating a skin condition, one of which can only be successfully formulated as a gel and the other of which can be formulated either as a gel or a cream. In comparing the two clinically, an explanatory mentality would favor simply comparing the two treatments administered as gels, while a pragmatic attitude would compare the first treatment as gel with the preferred version of the second treatment (ability to deliver as cream being considered an actual advantage of the second treatment rather than a pesky confound). Comparing Treatment A (substance 1 in a gel) and Treatment B (substance 2 in a cream) looks, from an explanatory mentality, like a clear case of confounding two possible causes; from a pragmatic mentality, it looks like a straightforward comparison of two actual treatments a patient might receive.

3.3 Outcomes

Schwartz and Lellouch point out that a pragmatic attitude prefers outcome measures that are close to what a patient and clinician are trying to accomplish with a course of treatment: a feeling of well-being, a remission of pain, a return to normal activity, an extension of life, or something similar. Some of these outcomes (death, for example) may be inconvenient or unethical in research, and others (anything involving patient self-assessment) have known validity problems. Explanatory clinical trials quite commonly use more convenient outcome measures that are known to correlate highly with the actual outcome of interest. For example, blood cholesterol levels are commonly used to assess preventive treatment for cardiovascular disease instead of tracking actual cardiovascular events such as heart attacks and strokes. The advantages of this kind of outcome measurement are obvious, but so are the limitations: A correlate of a disease may not be in any sense a cause of the disease, requiring (at some point) further evidence of effectiveness.

Schwartz and Lellouch were among the earliest to argue that explanatory trials would inevitably fall short of what would be needed to support clinical decision-making. Conclusions drawn from explanatory trials have superficial plausibility as means-end premises for practical reasoning about clinical decisions, as shown in Figure 3. Schwartz and Lellouch's arguments expose a serious threat to the validity of the conclusion: the means-end premise is plausible only if much too much is assumed about a demonstration of efficacy (specifically, that T's average efficacy justifies its use in every case, and that this is so irrespective of other possible treatments that may also be efficacious).

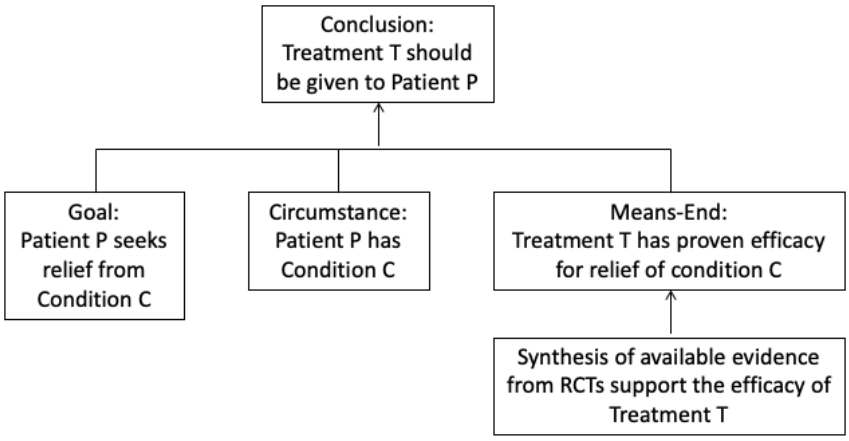

Figure 3 – Practical reasoning diagram, with Means-End premise drawn from explanatory RCTs.

Closing the gap between what is established by explanatory trials and what is needed for clinical care is a persistent theme in the many proposals for expanded use of pragmatic trials. A conservative approach to this is to simply add a Phase IV: Once efficacy is established through tightly controlled explanatory trials, go on to establishing effectiveness using more flexible and generalizable pragmatic trials. This is clearly not what Schwartz and Lellouch (1967) had in mind—but it is the obvious way to avoid starting over from scratch.

4. N-OF-1 TRIALS

N-of-1 trials, also called single patient trials, are RCTs that compare the effectiveness of two or more treatments on a single person. They were imported into medical science from experimental psychology (Guyatt, 2016). Their defining feature is that they produce meaningful conclusions for each individual patient; when repeated over many patients they may also support generalization, including generalizations about treatment variability of the kind shown earlier in Figure 2. Some advocates of evidence-based practice see N-of-1 trials as the highest form of evidence—as the top of an evidence pyramid of individual study designs (Guyatt, Rennie, Meade, Cook & American Medical Association, 2015, p. 11) or as one of the highest forms of evidence on treatment benefits and treatment harms, alongside systematic reviews (Howock et al., 2011).

A resurgence of interest in N-of-1 trials has been connected to their usefulness in clinical investigation (Guyatt, 2016), where N-of-1 trials offer potential benefits in comparison to other approaches. They are inexpensive compared to conventional RCTs enrolling many patients (Shamseer et al., 2015, p. 43). They can provide timely results to each individual patient, and a series of similar N-of-1 trials can be aggregated to estimate population level effects (Nikles et al., 2011, p. 479).

A particular advantage of N-of-1 trials is their closeness and relevance to clinical care, "making research more like practice and practice more like research" (Kravitz et al., 2014, pp. 7–8). By contrast, there are multiple limitations in applying RCTs to routine clinical care. One challenge is in generalizing from research populations to patient populations: "Patients recruited into RCTs differ from those who are eligible but not recruited in terms of age, sex, race, severity of disease, educational status, social class, and place of residence" (Rothwell, 2005, p. 86). In the past, researchers had more freedom to restrict eligibility for what they thought of as design reasons, so an additional complication is that older research may be based on narrow categories of patients such as white men between 20 and 40 years old. Such arbitrary restrictions on eligibility conditions are now more carefully scrutinized by oversight agencies. The past literature base of RCTs is particularly likely to exclude women, the elderly, and patients with comorbidities (Rothwell, 2005). Another challenge, as noted earlier, is that an average benefit for a treatment is no guarantee of consistent benefit at the individual level. While RCTs provide population-level estimates of the efficacy, they do not indicate which course of treatment is best for a given patient.

Answering these challenges, N-of-1 trials give the most direct evidence possible for what works best for the individual patient—at least when it is in fact possible for all options to be tried by the same patient. Not every condition is suitable for comparative N-of-1 trials. They are best applied to chronic conditions that are relatively stable, where the treatment has a fast onset (and ideally a short half-life; Nikles et al., 2011, p. 473). As presently conceived, N-of-1 trials are not suitable for areas such as surgery, where an irreversible treatment may be given, or critical care/emergency medicine, where a patient being stabilized cannot serve as their own control but rather should be compared with other patients receiving a different treatment.

CONCLUSION

In our prior work we have focused on new inference methods—new ways to draw conclusions that are either better than old ways of drawing conclusions, or that allow us to draw entirely new kinds

of conclusions. The central conceptual advance has been the idea of a warranting device—a proposed inference rule that generates conclusions whose quality is partly dependent on various kinds of assurances provided by the community that deploys the device. We are not prepared to say whether pragmatic trials and N-of-1 trials are new warranting devices, mainly because the work of building out these assurances has not yet been done—as it has been for RCTs and for Cochrane Reviews.

From this study, we learn that these new inference methods will often have limitations that are exposed only in argumentative practice. The normal output of RCT is a carefully qualified claim about the average effect of a medical treatment when given to patients like those observed. Despite their obvious epistemic strengths, RCTs commonly provide evidence for conclusions that are still an inferential step away from the clinically relevant decision: whether a particular treatment should be given to a particular patient. Further inference is required, beyond what RCT itself warrants, to get to the claim that the treatment should be given to a particular patient. That gap does not become apparent until the scientific result moves from the upstream context of explanatory research to the downstream context of practical reasoning about health care.

Both pragmatic trials and N-of-1 trials aim to address this inferential gap. None of the arguments in favor of pragmatic trials or N-of-1 trials are arguments against RCT. On the contrary, both are infused with the spirit of experimenting and committed to extending RCT further and faster. But as may be intuitively clear, both of these innovations have potential to change the way we look at RCT.

ACKNOWLEDGEMENTS: Thanks to: the James Lind Library for providing useful reference material on the history of evidence-based medicine; Kiel Gilleade for help with images; Neil Smalheiser for drawing our attention to empty systematic reviews; Tzu-Kun 'Esther' Hsiao for analysis of empty reviews; Cochrane for providing bulk access to data; and Quynh Pham and Jessica Ancker for their AMIA 2016 preconference "Beyond the RCT: Evaluating Innovations in the Learning Health System."

REFERENCES

Center for Drug Evaluation and Research. (2019, August 9). Investigational New Drug (IND) Application. Retrieved August 28, 2019, from U.S. Food and Drug Administration website: http://www.fda.gov/drugs/types-applications/investigational-new-drug-ind-application

De Angelis, C., Drazen, J. M., Frizelle, F. A., Haug, C., Hoey, J., Horton, R., Kotzin, S., Laine, C., Marusic, A., Overbeke, A. J. P. M., Schroeder, T. V., Sox, H. C., & Weyden, M. B. V. D. (2004). Clinical trial registration: A statement from the International Committee of Medical Journal Editors. *New England Journal of Medicine*, *351*(12), 1250–1251. https://doi.org/10.1056/NEJMe048225

Demotes-Mainard, J., & Kubiak, C. (2011). A European perspective – the European clinical research infrastructures network. *Annals of Oncology*, *22*(Suppl 7), vii44–vii49. https://doi.org/10.1093/annonc/mdr425

Guyatt, G. (2016). N of 1 randomized trials: A commentary. *Journal of Clinical Epidemiology*, *76*, 4–5. https://doi.org/10.1016/j.jclinepi.2015.09.020

Guyatt, G., Rennie, D., Meade, M., Cook, D., eds. (2015). *Users' guides to the medical literature: A manual for evidence-based clinical practice*. 3rd edition. American Medical Association & McGraw-Hill Education.

Howock, J., Chalmers, I., Glasziou, P., Greenhalgh, T., Heneghan, C., Liberati, A., Moschetti, I., Phillips, B., & Thornton, H. (2011). *Oxford Centre for Evidence-Based Medicine 2011 Levels of Evidence*. Retrieved December 16, 2019, from https://www.cebm.net/wp-content/uploads/2014/06/CEBM-Levels-of-Evidence-2.1.pdf

Jackson, S., & Schneider, J. (2018). Cochrane Review as a "warranting device" for reasoning about health. *Argumentation*, *32*(2), 241–272. https://doi.org/10.1007/s10503-017-9440-z

Kravitz R.L., Duan N., eds, & the DEcIDE Methods Center N-of-1 Guidance Panel (Duan N., Eslick I., Gabler N.B., Kaplan H.C., Kravitz R.L., Larson E.B., Pace W.D., Schmid C.H., Sim I., & Vohra S.). Design and Implementation of N-of-1 Trials: A User's Guide. AHRQ Publication No. 13(14)-EHC122-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2014. https://effectivehealthcare.ahrq.gov/topics/n-1-trials/research-2014-5/

Nikles, J., Mitchell, G. K., Schluter, P., Good, P., Hardy, J., Rowett, D., Shelby-James, T., Vohra, Su., & Currow, D. (2011). Aggregating single patient (n-of-1) trials in populations where recruitment and retention was difficult: The case of palliative care. *Journal of Clinical Epidemiology*, *64*(5), 471–480. https://doi.org/10.1016/j.jclinepi.2010.05.009

Office of the Commissioner. (2019, April 18). The Drug Development Process, Step 3: Clinical Research. Retrieved July 12, 2019, from U.S. Food and Drug Administration website: http://www.fda.gov/patients/drug-development-process/step-3-clinical-research

Rothwell, P. M. (2005). External validity of randomised controlled trials: "To whom do the results of this trial apply?" *The Lancet*, *365*(9453), 82–93. https://doi.org/10.1016/S0140-6736(04)17670-8

Schneider, J., & Jackson, S. (2018a). Innovations in reasoning about health: The case of the Randomized Controlled Trial. In B. Garssen, D. Godden, & G. R. Mitchell (Eds.), *Proceedings of the Ninth Conference of the International Society for the Study of Argumentation, Amsterdam, Netherlands July 2018* (pp. 1042–1050). Retrieved from http://cf.hum.uva.nl/issa/ISSA_2018_proceedings.pdf

Schneider, J., & Jackson, S. (2018b). Modeling the invention of a new inference rule: The case of 'Randomized Clinical Trial' as an argument scheme for medical science. *Argument & Computation*, *9*(2), 77–89. https://doi.org/10.3233/AAC-180036

Schwartz, D., & Lellouch, J. (1967). Explanatory and pragmatic attitudes in therapeutical trials. *Journal of Chronic Diseases*, *20*(8), 637–648.

Shamseer, L., Sampson, M., Bukutu, C., Schmid, C. H., Nikles, J., Tate, R., Johnston, B.C., Zucker D., Shadish, W.R., Kravitz, R., Guyatt, G., Altman, Douglas G., Moher, D., Vohra, S., & the CENT group. (2015). CONSORT extension for reporting N-of-1 trials (CENT) 2015: Explanation and elaboration. *BMJ*, *350*, h1793. https://doi.org/10.1136/bmj.h1793

Taichman, D. B., Sahni, P., Pinborg, A., Peiperl, L., Laine, C., James, A., Hong, S-T., Haileamlak, A., Gollogly, L., Godlee, F., Frizelle, F.A., Florenzano, F., Drazen, J.M., Bauchner, H. Baethge, C., & Backus, J. (2017). Data sharing statements for clinical trials: A requirement of the International Committee of Medical Journal Editors. *JAMA*, *317*(24), 2491–2492. https://doi.org/10.1001/jama.2017.6514

Thiers, F. A., Sinskey, A. J., & Berndt, E. R. (2008). Trends in the globalization of clinical trials. *Nature Reviews Drug Discovery*, *7*(1), 13–14. https://doi.org/10.1038/nrd2441

Toulmin, S. E. (1958). *The uses of argument*. Cambridge [England]: Cambridge University Press.

U.S. National Institutes of Health. (2014). *Notice of Revised NIH Definition of Clinical Trial* (No. NOT-OD-15-015). Retrieved from https://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-015.html

Yaffe, J., Montgomery, P., Hopewell, S., & Shepard, L. D. (2012). Empty reviews: A description and consideration of Cochrane systematic reviews with no included studies. *PLOS ONE*, *7*(5), e36626. https://doi.org/10.1371/journal.pone.0036626