Journal of Clinical Epidemiology 107 (2019) 71−76

Journal of Clinical Epidemiology

REVIEW

The shrinking scope of pragmatic trials: a methodological reflection on their domain of applicability

Ray Pawson*

Emeritus Professor in Social Research Methodology, School of Sociology and Social Policy, University of Leeds, Leeds LS2 9JT, UK

Accepted 23 November 2018; Published online 28 November 2018

1. The shrinking scope of pragmatic trials

The appliance of science is always challenging. How can empirical findings won in closed systems, with rigid research designs, specially constructed technical apparatuses, and erudite analytic techniques be used as a guide to action in routine, everyday practice? In the clinical realm the key attempt to unravel this puzzle has been undertaken in the name of pragmatic trials, a 50-year-old mission, much of which has been played out in the pages of this journal.

The gap to be bridged has been described a thousand times. Sitting atop, the evidence hierarchy is the placebo-controlled randomized controlled trial, which thanks to application of further controls for selection bias, performance bias, adherence bias, detection bias, attrition bias, reporting bias, and so on permits the inference that the intervention and the intervention alone has causal efficacy. But what experimentalists regard as bias is likely to be perceived as the norm in routine, everyday conditions for the application of that intervention. And what practitioners want to know is how the intervention will fare in their world when they confront specific patients with their own comorbidities, when those patients comply unevenly with and sometimes drop their medications, when staff advise upon and implement treatments inconsistently, when those treatments are rationed, when disparate stakeholders seek different, short- and long-term outcomes, and so on.

The preferred solution is easy to utter but difficult to realize, namely ‘‘make the clinical trial more responsive to real-world conditions.’’ A vast literature has gathered in response to this cause, prompted by Schwartz and Lellouch’s founding contribution [1]. Methodological interest has quickened in recent years, signified by the sustained effort to formalize and operationalize the difference between explanatory and pragmatic trials. Whole research teams have dedicated themselves to this task, most notably under the auspices of the ‘‘PRECIS’’ tool [2,3] with its famous spokes-hub-and-rim, bicycle-wheel diagrams. The current tool identifies nine dimensions (the spokes) with which to differentiate explanatory and pragmatic trials. Researchers assess any particular or proposed trial by applying a score (Likert scale 1−5) on each dimension, with more explanatory designs being placed near the hub and more pragmatic trials located nearer the rim. The overall configuration (relatively explanatory or relatively pragmatic) is then signified by a perambulatory line, which connects up the selected scores. The anatomy of any trial is thus laid bare, and users can assess to extent to which their workaday concerns are addressed.

All this effort has been discharged with the objective of improving the ‘‘generalizability’’ or ‘‘external validity’’ or ‘‘scope conditions’’ of clinical trials. This article provides a constructive critique of that claim. As we shall see, expectations on these issues have changed considerably over the years, and it is appropriate to begin at the beginning. The initial ambition for broadening the scope of clinical trials resided in the fact that pragmatic trials operated in ‘‘real-world’’ conditions as opposed to the ‘‘artificial’’ conditions as manipulated in an efficacy trial. Scores of authors, especially in the original discussions, used the ‘‘real-world’’ motif to define the domain of a pragmatic randomised controlled trial (PRCT). And it was this new-found ability to mimic everyday clinical practice that was said to render pragmatic trials more generalizable. Ware and Hammel, for instance, make the following ringing declaration: ‘‘Pragmatic trials are designed to study real-world practice and, therefore, represent less-perfect experiments than efficacy trials: they sacrifice internal validity to achieve generalizability’’ [4].

The present contribution interrogates this proposition and does so from the perspective of the methodological literature on inferential logic. All the applied sciences face an equivalent challenge of generalizing from restricted settings, prototypes, samples, case studies, simulations, and demonstration projects. This dilemma, writ large, has led to prolonged scrutiny of the logical foundations of external

Conflict of interest: The author declares that there is no conflict of interest.

* Corresponding author.
E-mail address: r.d.pawson@leeds.ac.uk.

https://doi.org/10.1016/j.jclinepi.2018.11.020

What is new?

Explanatory trials contrived to maximize therapeutic benefit are giving way to pragmatic designs that are more responsive to the quotidian concerns of practitioners and patients. The article revisits one of the original objectives of that quest, namely to enhance the "generalizability" of trial findings.

Key findings

- The issue of generalizabilty has received considerable attention in the philosophy of science literature but remains a contested concept. Three venerable interpretations are distinguished—simple generalization, extensional generalization, and applicability. The pragmatic trial literature has not been well informed by these discussions and has veered between the three meanings.
- The original claims that pragmatic trials are typical (simple generalization) and inform a range of local and distant settings (extensional generalization) have retrenched to the view that their results only relate to specific conditions, which match those operating in a particular pragmatic trial (applicability).

What this adds to what was known?

- This move to narrow the domain of pragmatic trials introduces an unforeseen convergence with the highly restricted scope of explanatory trials. Both generate case-specific findings. Any difference in results between the two trial modes is simply an indication that the therapy works in some situations and not others.

What is the implication and what should change now?

- Generalization cannot be achieved in the single trial—be it explanatory or pragmatic. It requires a multicase, multimethod approach building an understanding of the biological and behavioral mechanisms of action that give rise to heterogeneous treatment outcomes.

validity, and it turns out that high road from the particular to the general is now understood to take many different pathways, namely—utility, simple and extensional generalization, abduction, retroduction, applicability, transportability, transferability, predictability and so on [5–8]. The pragmatic trial literature has not been well informed by these finer distinctions, and the study examines the claims made for the widening scope of pragmatic trials against the three of these concepts, namely—simple generalization, extensional generalization, and applicability.

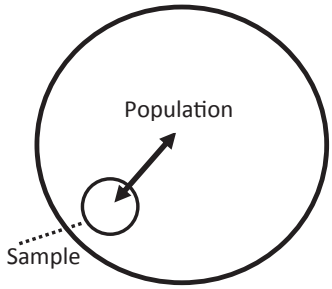

Fig. 1. Simple generalization: extrapolation on the same dimension.

1.1. Simple generalization

Simple generalization has its roots in sampling theory. On this model, the central concern is warranting how findings from a sample can speak for the whole population from which it is drawn. A typical example from the social realm might concern the perils of opinion polling and whether the voting preferences of the chosen sample will reflect accurately the views of the total population as revealed in an election. The crucial design issue thus concerns the "representativeness" of those investigated [5]. Obeying a range of statistical protocols on selection methods, sampling frames, sampling fractions, and confidence intervals ensures that the chosen sample has adequate statistical power to justify the claim that it indeed "represents" or is "typical" of the population as a whole. The key point is that this model of external validity rests on extrapolations *on the same dimension* (e.g., some voters to all voters). This form of generalization is illustrated in Figure 1 in which the sample is embedded in the wider population.

Is this the understanding of generalization assumed in a pragmatic trial? In Sackett's famous "primer" on the two forms of clinical trials [9], he does indeed equate pragmatic trials with typicality, wanting them to answer this question: "Does this treatment improve patient-important outcomes when applied by typical clinicians to typical patients" (underlining in the original). The appropriate course of action seems straightforward enough—for example, a pragmatic trial should deliberately widen the patient population from that investigated in an efficacy trial to make it more "representative."

But does such a maneuver make the ensuing findings widely applicable? Clearly not—for the selection of individual patients entering the pragmatic trial is but one of its constituents. As PRECIS profiling teaches us, it is the intervention as a whole that makes up the pragmatic trial. And that intervention is a complex biological, technical, and social system, comprising patients undergoing the intervention, practitioners implementing the intervention,

researchers evaluating the intervention, institutions administering the intervention, providers funding the intervention, regulators approving the intervention, and health services rationing the intervention. In short, the ability to generalize from a trial is not confined to the single dimension (e.g., some patients to all patients). All the above features condition trial outcomes. Pragmatic trials are complex interventions and to generalize from a pragmatic trial is to extrapolate from a complex system to other complex systems. And this is not within the compass of simple generalization. There is no such thing as a representative place in an evolving, multidimensional space.

1.2. Extensional generalization

This brings us to the idea of "extensional generalization" in which the basis of the claim is that "situations of *other* kinds will be like those observed kinds" [5]. Burchett et al. provide another definition better to capture the intended scope: "external validity—the likelihood that a study's findings could be generalized to other (unspecified and more general) samples and settings" [10]. This claim can be located throughout the pragmatic trial literature with authors repeatedly using phrases such as "usual care" [4], "routine clinical and healthcare settings" [11], and "usual community of users" [12] to indicate the extensive target domain.

Figure 2 describes this "one-to-many" objective. At the bottom, represented (perhaps overdramatically) by the explosion symbol, we have the complex intervention investigated in the pragmatic trial. Above lies the real world, represented by six token interventions in six different contexts in which the "same" intervention is applied. Two of the symbols are dashed to represent the idea that some of these further settings are "unspecified and more general." The arrowed lines capture the expectation that the findings from the pragmatic trial can be applied to these other routine settings.

Is this interpretation of the external validity of pragmatic trial justified? Clearly not—for exactly the same objection applies. The pragmatic trial investigates one incarnation of an intervention applied to a particular range of patients by a particular set of practitioners in a particular institutional setting in a particular health service. There is no reason whatsoever to suppose that the "same" intervention applied differentially to a different range of patients by a different range of stakeholders in different institutional settings should carry the same net effect. Extensional generalization depends on the capacity for research to establish "likeness" between one system and another and another [13]. No such warrant is available without detailed examination of each of those other settings. The real world is a plurality of any number of "real worlds." Folding them together under abstractions such as "usual care," "routine clinical settings," "community of users," and so on cannot be justified.

And with this dawning realization the cherished conviction that pragmatic trials have widespread, extensional generalizability has died. This retreat is most clearly signaled in a recent exchange between Dekkers et al. [14] and Zwarenstein et al. [12]. The former takes umbrage with the much-quoted idea that the PRECIS tool identifies those pragmatic trials that will inform "a range of local and distant settings." The latter team protest that this was never their intention and specify the precise ambitions for usage of the tool as follows: "Our conception of PRECIS-2 is that it is to be used by trialists to design a trial whose results are applicable to a context in which they, the trialists, are intending the results to be used." In other words the goal is to extrapolate findings from a specific pragmatic trial to the specific application in which the same contextual conditions apply. The bounteous notion of extensional validity gives way to the parsimonious concept of "applicability."

1.3. Applicability

We have followed the significant retreat from the original, rose-tinted expectation that pragmatic trials would enjoy widespread use across real-world practice to the current, hard-headed view that particular pragmatic trials will only find application in corresponding clinical settings.

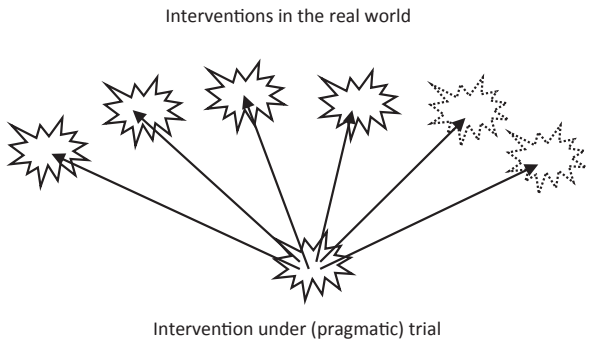

Fig. 2. Extensional generalization: system-to-systems extrapolation.

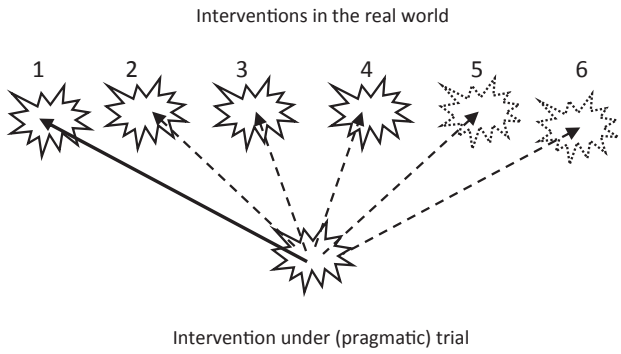

Fig. 3. Application: case-to-case extrapolation.

What are the implications? The present one-to-one view of applicability is illustrated in Figure 3.

It is now understood that a pragmatic trial will find ready application in those clinical settings where the routine discharge of the treatment resembles most closely the design features as captured in, say, the nine key PRECIS dimensions of that trial. The capacity for utilization is depicted, once again, by arrow 1 connecting that trial and that clinic. However, it is now clearly acknowledged that the same intervention will find use in many other settings in which different external ''threats'' occur and in which different PRECIS profiles will therefore apply. The findings from the original pragmatic trial cannot, therefore, be extrapolated across these further scenarios (2−6), and accordingly the connecting arrows become dashed.

Although the logic here is impeccable, it generates two unintended methodological consequences. The first is pictured directly in Figure 3. Case 1 exemplifies a concrete application of a pragmatic trial. Research user 1 scrutinizes the configuration of contextual factors encompassing a given pragmatic trial and deems them close enough to those operating in her institution; she implements the treatment in the same manner, and evidence utilization becomes a reality. But what of users 2, 3, 4, 5, 6 … N? They decipher that pragmatic trial 1's characteristics such as patient eligibility, practitioner expertise, practitioner adherence, comparator flexibility, outcomes monitored, follow-up intensity, and so on are not the same as on their patch. Where do they turn? Do they have to await the results of tailor-made, pragmatic trials that match the characteristics of implementations 2, 3, 4, 5, 6 … N? Given the substantial expense of mounting trials and the trundling pace of trial results and the competing priorities facing trial funders, it is most unlikely that the necessary mosaic of pragmatic trails could ever be assembled. Clearly, they need to look elsewhere for a different kind of evidence.

The second consequence of narrowing the domain of applicability is to introduce an unforeseen convergence between the shrinking scope of pragmatic trials and that of explanatory trials. It has always been acknowledged that the latter trials, also called efficacy trials, generate findings that apply in highly restricted conditions. The classic problem is that these conditions are often regarded as ''artificial'' or ''favorable,'' hence the move to locate the trial in the rugged, real world. But now we learn that the findings from pragmatic only apply in restricted settings and conditions. The consequence, first spotted by Kent and Kitsios [15], is that neither form of trial should be considered more informative; neither indicates the ''true effect'' of a treatment. Any difference in results between the two trial modes is simply an indication that the therapy works in some situations and not others. A lesson from physical science is used to hammer home this message:

''Determining the 'true' treatment effect of a given therapy is a bit like determining the 'true' weight of a liter of water. Those who answer that a liter of water weighs a kilogram are either assuming an implicit 'on planet earth, at sea level, at 4° Celsius,' or confusing the intrinsic property of mass with the extrinsic property of weight. Similar to weight, treatment effect is an extrinsic property, emerging only through an interaction between the intervention, the patient, and the circumstances in which it is being measured. Adjust the context and a different effect emerges, just as a liter of water weighs a little over one-third of a kilogram on Mars'' [15].

1.4. Conclusion: stretching the scope of pragmatic inquiry

Fifty years on, it turns out that pragmatic trials attenuate utility. They only apply in clinical situations that match, in so far as this is possible, the patients, the practitioners, the implementation regime, the therapeutic environment, and the outcome portfolio as examined in a particular trial. The PRECIS exercise describes these scope conditions with previously unsurpassed precision but then foregoes widespread application. So how might we stretch the scope of pragmatic inquiry? The core requirement is a fundamental shift from the narrow shoulders of individual trials to a model of explanation building based on the entire body of available evidence. It is impossible to provide a methodological blueprint in the space of a brief conclusion, but some key principles are sketched in the following paragraphs.

We have seen that a pragmatic trial represents just one of the many ways and many circumstances in which treatment may be exercised. An important corollary follows—the findings of that pragmatic randomised controlled trial represent just one of the many potential outcomes of that treatment. With this predicament, we enter discussion of one the great challenges of evidence-based medicine (EBM), namely the ''heterogeneity of treatment effects'' (HTE). Common sense holds that treatment effects are not the same for every patient. The same, nonrandom variation occurs within all RCTs. In a series of articles, Kent et al. [16−18] demonstrate the significant extent of HTE, even in efficacy trials with tightly controlled inclusion and exclusion criteria. Within that trial population, there will still be residual variation (albeit more limited than at posttrial, postapproval, etc.) in terms of age, sex, disease etiology, patient history, concomitant exposures, competing risks, and genetic background. The implications for EBM are considerable:

When HTE is present, the modest benefit ascribed to many treatments in clinical trials can be misleading because modest average effects may reflect a mixture of substantial benefits for some, little benefit for many, and harm for a few [19].

The crucial implication for this article is that pragmatic trials amplify HTE *by design*, a point first made in a grossly overlooked paper by Segal et al. [20]. Consider the patient

eligibly spoke of the PRECIS graphic. A pragmatic trial, with its generous eligibility criterion, has the effect of changing the risk profile, modifying the subgroups of potential responders and nonresponders to treatment. Proceeding around the other domains, pragmatic randomised controlled trials complicate the pattern of patient compliance, expand differences in practitioner expertise, distend flexibility of implementation, and so on, all driving up HTE.

Conclusion? Instead of seeking the chimera of the ''true effect'' or settling for the ambiguity of the ''net effect'' or striving for the nonexistent ''typical effect,'' EBM should be organized around indentifying, understanding, and exploiting ''heterogeneous effects.'' There is no particular design to achieve this objective; its realization depends on creating *series* or *sequences* of studies, using different methods, focusing on different physiologies, patients, and practitioners. Two, very brief and highly contrasting, examples illustrate the broad strategy.

Tumor heterogeneity, ranging from spontaneous regression to relentless progression, is the great bugbear of cancer treatment trials. Research continues apace and, even if an established biomarker is used as the basis for patient selection, trials still generate HTE [21]. The solution is to go back to the drawing board and into the laboratory, to devise more sophisticated theories of how ''genetic and epigenetic heterogeneity affects tumor evolution'' [22], to incorporate these ideas into a revised treatment, and then to return these hypotheses to the field for empirical testing. The perpetual goal is to gradually increase the proportion of suitable patients receiving appropriate treatments, and this is realized in an ever-lasting research process. Such research cycles may occur in the long or short term, a recent example of the latter being the ''biomarker-driven adaptive population enrichment design,'' which makes ''prospectively planned change to the trial on the basis of an analysis of accumulating data from the trial itself'' [23]. No details are provided here. What matters is the principle, namely, that it is the adaptive process of treatment, testing, recognizing error, proposing refined solutions and retesting, which holds the key to generalizable knowledge.

Reforms in health service delivery are notoriously difficult to achieve and evaluate. They are bedeviled by diverse forms of practitioner resistance and by assorted institutional inflexibilities, of which there is no single pattern, and no single trial can decipher. The solution is for implementation science to focus on the program theory and to test it in multiple, within-site and between-sites inquiries. Benzer et al. [24] undertook such a study on the attempted integration of primary care and mental health services in Veterans Affairs Centers in the United States. The underlying theory was that co-location of the services would improve and quicken access. Outcomes were disappointingly mixed. The research team explains this heterogeneity by mapping the divergent professional views (within sites) and varying institutional barriers (between sites). There is no claim here that there is a typical intervention and a typical response. But there are some more subtle patterns, which stem from the respective power and the appetite for collaboration of various stakeholders. Knowledge of how these configurational patterns operate is the key to future attempts at service reform.

All interventions have a base in theory, whether that theory is biological, genetic, physiological, psychological, or social. All these theories are fallible and will hold only, in particular, applications for particular patients in particular circumstances. Investigative series and sequences are the keys to identifying, understanding, and exploiting such heterogeneous effects. The approach sketched here might conceivably include findings from a particular pragmatic trial, but only if they are deeply embedded in a wider multidisciplinary, multimethod evidence base.

There is nothing as pragmatic as a good, robustly tested theory.

References

[1] Schwartz D, Lelloch J. Explanatory and pragmatic attitudes in therapeutic trials. J Chronic Dis 1967;20:637−48.

[2] Thorpe K, Zwarenstein M, Oxman A, Treweek S, Furberg C, Altman D, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. J Clin Epidemiol 2009;62:464−75.

[3] Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe K, Zwarenstein M, et al. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ 2015;350:h2147.

[4] Ware J, Hamel M. Pragmatic trials — guides to better patient care? N Engl J Med 2011;364:1685−7.

[5] Kaplan A. The Conduct of Inquiry. San Francisco: Chandler Publishing Company; 1964.

[6] Campbell D. Relabeling internal and external validity. In: Campbell D, Russo M, editors. Social Experimentation. Thousand Oaks, CA: Sage; 1999.

[7] Danermark B, Ekstrom M, Jakobsen L, Karlsson J. Generalization, scientific inference and models for an explanatory social science in B Danermark et al. Abingdon: *Explaining Society* Routledge; 2002.

[8] Murad M, Katabi A, Benkhadra R, Montori V. External validity, generalisabilty, applicability and directness: a brief primer. BMJ Evid Based Med 2018;23:17−9.

[9] Sackett D. Explanatory and pragmatic clinical trials: a primer and application to a recent asthma trial. Pol Arch Med Wewn 2011; 121(7−8):259−63.

[10] Burchett H, Umoquit M, Dobrow M. How do we know when research from one setting can be useful in another? A review of external validity, applicability and transferability frameworks. J Health Serv Res Policy 2011;16(4):238−44.

[11] Sedgwick P. Explanatory versus pragmatic trials. BMJ 2014;349: g6694.

[12] Zwarenstein M, Treweek S, Loudon K. PRECIS-2 helps researchers design more applicable RCTs while CONSORT Extension for Pragmatic Trials helps knowledge users decide whether to apply them. J Clin Epidemiol 2017;84:27−9.

[13] Goodman N. Likeness. In: Goodman N, editor. Problems and Projects. Indianapolis: Bobbs-Merrill; 1972.

[14] Dekkers O, Bossuyt P, Vandenbrouke J. How trial results are intended to be used: is PRECIS-2 a step forward? J Clin Epidemiol 2017;84:25−6.

[15] Kent D, Kitsios G. Against pragmatism: on efficacy, effectiveness and the real world. Trials 2009;10:48.
[16] Kent D, Hayward R. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. JAMA 2007;298:1209−12.
[17] Kent D, Hayward R. When averages hide individual differences in clinical trials: analyzing the results of clinical trials to expose individual patients' risks might help doctors make better treatment decisions. Am Sci 2007;95(1):60−9.
[18] Kent D, Rothwell P, Ioannidis J, Altman D, Hayward R. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. Trials 2010;11/1:85.
[19] Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment-effects, and the trouble with averages. Milbank Q 2004;82(4):661−87.
[20] Segal J, Weiss C, Varadhan R. Understanding heterogeneity of treatment effects in pragmatic trials: a white paper on behalf of the center for medical technology and policy 2011. Available at: http://www.cmtpnet.org/docs/resources/Segal-Heterogeneity-in-Pragmatic-Trials.pdf. Accessed December 17, 2018.
[21] Maris J. The biologic basis for neuroblastoma heterogeneity and risk stratification. Curr Opin Pediatr 2005;17(1):7−13. https://doi.org/10.1097/01.mop.0000150631.60571.89.
[22] Alizadeh A. Toward understanding and exploiting tumor heterogeneity. Nat Med 2015;846−53.
[23] Bhatt D, Metha C. Adaptive designs for clinical trials. N Engl J Med 2016;375:64−74.
[24] Benzer J, Beehler S, Cramer I, Mohr D, Charns M, Burgess J. Between and within-site variation in qualitative implementation research. Implement Sci 2013;8:4.