ELSEVIER

Journal of Clinical Epidemiology 63 (2010) 980–982

Journal of Clinical Epidemiology

MISSED LEARNING OPPORTUNITIES TO REFINE THE DESIGN OF STUDIES

Learning from the past: refining the way we study treatments

Wolfgang Viechtbauer*

Department of Methodology and Statistics, School for Public Health and Primary Care (CAPHRI), Maastricht University, Maastricht, The Netherlands

Accepted 12 April 2010

In light of economical and ethical considerations, methods and approaches that help researchers use information from previous research in the design and reporting of future studies can be expected to receive increasing attention in the coming years. The present issue of the *Journal of Clinical Epidemiology* includes two timely articles that pertain to this idea in various ways. Before briefly commenting on these two articles, it may be useful to put this work in a larger context.

Introductory textbooks on research methods in the health, social, and medical sciences usually describe a cyclical process by which knowledge is accumulated and refined by means of empirical research [1,2]. In its basic form, the process involves the formulation of theories by means of inductive reasoning (i.e., forming conjectures about the likely relationships between variables and the reasons for them), the derivation of suitable hypotheses from these theories by deduction (i.e., making predictions about expected observations assuming that the underlying theories are true), the testing of these hypotheses in appropriately conducted studies (entailing the possibility of actually falsifying the stated hypotheses), with the conclusions from those studies then leading to revision and refinement of the corresponding theories. The accumulation of knowledge by means of replication and the refinement of theories over time can, thus, be regarded as cornerstones of scientific research. Formal descriptions of this process can be found, for example, in the works of Popper [3] and Hempel [4].

Accordingly, careful use of the information obtained from previous studies is an integral aspect of conducting research. In fact, *The Lancet* [5] and several other journals (including the *Journal of the American Medical Association* and the *Archives of Internal Medicine*) now require that authors demonstrate the need for a new trial in light of the existing evidence basis and set its results in the context of the previous studies. This can usually be accomplished by considering the results from a relevant meta-analysis (that demonstrates the need for additional studies) and by updating its results with the findings from the new trial.

Although this iterative research process and the updated journal guidelines tend to emphasize how the *substantive* conclusions and findings from research lead to increased understanding of the phenomenon under study, relatively little has been written on how the *design* of future studies should be informed by existing research [6]. Therefore, although we are constantly striving to refine our theories about the phenomenon itself (e.g., through increased knowledge of how large an effect is on average, how variable the effect is, or for whom and under which circumstances the effect is larger or smaller), we are not necessarily refining *how* we study the phenomenon. Similarly, systematic reviews and meta-analyses of the literature tend to emphasize what has been found rather than what has been studied and how [7].

Based on a sample of 27 trials published in the *Journal of the American Medical Association* and the *Archives of Internal Medicine*, Goudie et al. [8] document the extent to which the motivation and the design of trials have been informed through previous studies and discuss how trialists can take advantage of the existing knowledge base when designing and reporting the results of their studies. Like earlier surveys of this type (see article for references), the findings suggest that previous studies (or relevant systematic reviews) are not considered to the fullest extent possible when designing or reporting the results of new trials.

Not surprisingly, one of the design elements that is most frequently informed by either previous studies or systematic reviews thereof is the sample size for a new trial. Both Goudie et al. [8] and Borm et al. [9] discuss methods that can help trialists in their sample size calculations. Goudie et al. [8] illustrate how to use a novel method for determining a suitable sample size when the ultimate goal is to update an existing meta-analysis with the findings of the new trial, whereas Borm et al. [9] describe a simple method for determining the sample size of a new trial to achieve a certain power in its own right when only the sample size and the *P*-value from a ‘‘reference trial’’ are known. Both

* Corresponding author. Department of Methodology and Statistics, School for Public Health and Primary Care (CAPHRI), Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands. Tel.: +31-43-388-2277; fax: +31-43-361-8388.

E-mail address: wolfgang.viechtbauer@stat.unimaas.nl

0895-4356/$ - see front matter
doi: 10.1016/j.jclinepi.2010.04.004

of these approaches should be carefully considered by trialists, because one automatically leads to a consideration of the totality of the existing evidence when designing studies, whereas the other helps to simplify potentially complicated sample size calculations for individual trials.

Interestingly, these two approaches draw attention to a potential dilemma that an increased emphasis on updated meta-analyses may entail. In particular, there is a conflict of interest between researchers conducting individual trials (for whom, generally speaking, a single decisive ''mega-trial'' may provide more valid and informative results than several smaller studies) and meta-analysts (for whom several smaller trials may be more useful for estimating the amount of heterogeneity and delineating the generalizability of the findings than one large study).

In an ideal case, the design of a set of trials and a resulting meta-analysis incorporating those trials should go hand in hand, a model that has occasionally been put into practice in some prospectively planned meta-analyses [10,11]. Sample size considerations can then be made cooperatively and the ''credit'' for the individual trials and the resulting meta-analysis can be automatically shared among members of such collaborations. The increased tendency, for example, in the European Union, to start international research projects may, in fact, be an incentive to plan some prospective meta-analyses. However, such collaborations are likely to remain exceptions in practice, leaving sample size considerations with the individual trialist planning a single study. It remains to be seen how editors, reviewers, and granting agencies will then react to (proposals for) one or multiple studies that may be underpowered on their own but contribute sufficient evidence to power an updated meta-analysis.

A related issue is the motivation for conducting a new trial in the first place. Deciding at which point further research should be deemed unnecessary cannot and should not be made by reference to *P*-values and power alone. In an extreme sense, a single sufficiently powered study with a significant outcome would then put a stop to any further research on a particular treatment. Given that the tendency to only publish significant findings is especially great when commencing research into new treatments [12], one has to be very careful with halting additional studies when the results from an early meta-analysis suggest a significant treatment effect. Moreover, heterogeneity in the treatment effectiveness may only become apparent after a sufficient number of trials have been carried out. Cases in which the finding of an effective treatment (as determined by an early meta-analysis) was later turned over by subsequent trials (e.g., the third example in the article by Goudie et al.) further highlight the difficulty in choosing an appropriate stopping point.

The replication paradox discussed by Borm et al. [9] also helps to put this issue into perspective. In particular, a meta-analysis can, in essence, be regarded as a sort of multicenter trial for which we can also determine the probability of a statistically significant replication. Therefore, just as for an individual study, only half of all meta-analyses yielding an overall effect with $P = 0.05$ will turn out to be significant again should one attempt to replicate their findings by repeating each of the component studies. Moreover, this holds true regardless of the number and size of the individual studies included in the initial meta-analyses. In that sense, *P*-values near 0.05 cannot be regarded as robust evidence, even when these are based on dozens of trials. Supplementing the results from a null hypothesis test with an estimate of the probability that a replication study would turn out to be significant may, in fact, be one way to emphasize the fact that *P*-values do not directly reflect the probability of a successful replication [13,14].

Finally, the focus on sample size considerations should not draw our attention away from other design issues that can and should be informed by previous research, such as the choice of outcome measures, the number and timing of the measurement occasions, the choice of confounders and/or effect modifiers to measure and include in the analysis, the choice of statistical model and analysis technique, and the inclusion/exclusion criteria for the participants. For example, using outcome measures with less measurement error (and hence, *ceteris paribus*, higher reliability) leads to more powerful tests [15] and/or smaller sample sizes [16]. Choosing three or more measurement occasions wastes resources if trends are, in fact, linear [17]. Although some aspects are routinely taken into account when designing new studies, it is clear that there is still much room for improving and formalizing the process by which we use previous research when we study the effects of treatments.

In conclusion, properly motivating a study in light of the existing evidence basis and setting its results in the context of previous research is one challenge for trialists. Another challenge is to improve and refine the design of new studies based on what has been done in the past. Meta-analysts are in a unique position of not only being able to describe what has been found but can also provide an overview of what has been studied and how (i.e., what has been called the ''study space'' [7]). Turning this information into practical guidelines for refining the way we study treatments should become a primary objective for methodologists in the future.

References

[1] Bordens KS, Abbott BB. Research design and methods: a process approach. 7th ed. New York, NY: McGraw-Hill; 2007.
[2] Crosby RA, DiClemente RJ, Salazar LF. Research methods in health promotion. San Francisco, CA: Wiley; 2006.
[3] Popper K. The logic of scientific discovery. London, UK: Hutchinson; 1959.
[4] Hempel CG. Aspects of scientific explanation and other essays in the philosophy of science. New York, NY: Free Press; 1965.
[5] Young C, Horton R. Putting clinical trials into context. Lancet 2005;366:107–8.
[6] Sutton AJ, Cooper NJ, Jones DR. Evidence synthesis as the key to more coherent and efficient research. BMC Med Res Methodol 2009;9.

[7] Malpass RS, Tredoux CG, Compo NS, McQuiston-Surrett D, MacLin OH, Zimmerman LA, et al. Study space analysis for policy development. Appl Cogn Psychol 2008;22:789−801.
[8] Goudie CA, Sutton AJ, Jones DR, Donald A. Empirical assessment suggests that existing evidence could be used more fully in designing randomized controlled trials. J Clin Epidemiol 2010;63:983−91.
[9] Borm GF, Bloem BR, Munneke M, Teerenstra S. A simple method for calculating power based on a prior trial. J Clin Epidemiol 2010; 63:992−7.
[10] Schechtman KB, Ory MG, the FICSIT Group. The effects of exercise on the quality of life of frail older adults: a preplanned meta-analysis of the FICSIT trials. Ann Behav Med 2001;23:186−97.
[11] Simes RJ. Prospective meta-analysis of cholesterol-lowering studies: the prospective pravastatin pooling (PPP) project and the cholesterol treatment trialists (CTT) collaboration. Am J Cardiol 1995;76:122C−6C.
[12] Ioannidis JPA. Why most discovered true associations are inflated. Epidemiology 2008;19:640−8.
[13] Greenwald AG, Gonzalez R, Harris RJ, Guthrie D. Effect sizes and p values: what should be reported and what should be replicated? Psychophysiology 1996;33:175−83.
[14] Posavac EJ. Using p values to estimate the probability of a statistically significant replication. Understanding Stat 2002;1:101−12.
[15] Williams RH, Zimmerman DW, Zumbo BD. Impact of measurement error on statistical power: review of an old paradox. J Exp Educ 1995;63:363−70.
[16] Perkins DO, Wyatt RJ, Bartko JJ. Penny-wise and pound-foolish: the impact of measurement error on sample size requirements in clinical trials. Biol Psychiatry 2000;47:762−6.
[17] Berger MPF, Wong WK. Applied optimal designs. Chichester, UK: Wiley; 2005.