Effectiveness in practice-based research: Looking for alternatives to the randomized controlled trial (RCT)

Tavecchio, L.

Published in: Methodological Review of Applied Research

Citation for published version (APA):
Effectiveness in practice-based research: Looking for alternatives to the randomized controlled trial (RCT)\(^4\)

Louis Tavecchio  
Hogeschool van Amsterdam (HvA), Amsterdam, The Netherlands,  
l.w.c.tavecchio@uva.nl

Abstract  
Over the last decade, the status of the randomized controlled trial (RCT), hallmark of evidence-based medicine (research), has been growing strongly in general practice, social work and public health. But this type of research is only practicable under strictly controlled and well-defined settings and conditions. In addition, persons are randomly assigned to experimental and control groups. However, in the daily practice of social work and public health professionals, this type of research is hardly feasible. Moreover, due to political, legal or ethical considerations, random selection or assignment will often be impossible. It is argued that the external validity of research outcomes will suffer from many of the features of the control in classic explanatory RCTs, not only for population-level interventions, but these outcomes will also be only marginally useful for deriving guidelines for professional practices and activities. Another big disadvantage of the RCT supremacy is the gap – i.e., loss of time and information - between research and practice, as illustrated by the ‘pipeline fallacy’ metaphor. In short, RCT may be the gold standard of study design in (bio) medical research, but it is often not feasible, ethical, or generalizable in the context of practice-based or practically-oriented research. In recent years, the appreciation and use of alternative research designs and data sources - that will overcome the disadvantages and limitations of RCTs - have increased. The article concludes with the description of alternatives to the classic explanatory RCT, i.e., methodological remedies for practice-based research with less emphasis on internal and more emphasis on external validity.

Summary of the reviews  
Received: 2014-09-05.  
Reviewer A 2014-09-12. The author’s view that RCT’s focus too much on internal validity at the expense of internal validity is clear, but not new. Apart from arguments by others, no new arguments are provided. Nowhere in the paper is discussed why other methods would lead to more or faster impact. They may have drawbacks as well. The methodological issues are discussed for the RCT, but not for the alternative methods. An example of practice-based research could strengthen the paper. The figures include details that are not explained. Recommendation: Resubmit for review.  
Reviewer B 2014-12-11. The organization of the paper is insufficient. Figures 2 and 3 concern issues that are not related to the design of studies. Do the authors wish to concentrate on questions on efficacy and effectiveness or on other issues related to the implementation and acceptability? Do they wish to discuss the design of studies or the ways in which research findings are disseminated, integrated and converted into guidelines? Should biomedical research be opposed to practice oriented research? It is unclear from the alternative study designs how they could address the issue of external validity. The writings of Pawson on realistic methods provide an example of how the issue of external validity could be addressed. The authors seem to neglect that a broad field of implementation research complements efficacy research. The authors should specify how for instance observational research can assess the effectiveness of interventions with some reasonable degree of internal validity, e.g., evaluations of natural policy experiments. It is unclear what the examples illustrate. The

\(^4\) This article is based on chapter 5 of Tavecchio & Gerrebrands (2012) and is an adaptation of a recently published Dutch article (Tavecchio, 2014).
example of Poortinga seems to fit poorly in a paper on effectiveness. The section on Pragmatic RCT does not clearly specify how this differs from Explanatory RCT in design. Recommendation: Decline submission.

Editor decision 2014-12-11: Resubmit for review.
Author response 2015-1-6: The author indicated that he could not provide for a major revision, due to lack of time.
Editor re-decision 2015-1-10: As the paper concerns an important subject, e.g., the applicability of research methods in practice based research, he proposed to the author to accept the paper in its current form, with the summary of the reviews, rejoinders from at least two critics and the opportunity for a reply. The author accepted these conditions. Both reviewers refused to extend their reviews to a publishable rejoinder. The editor decided to write a rejoinder. Furthermore, Piet Verschuren wrote an additional comment.

**Introduction**

In the age of evidence-based practice (EBP), few people seem to doubt the importance of utilizing research data for developing, implementing and evaluating policy measures and plans, interventions and methods development. Professionals in the fields of social work, welfare and public health will have to adopt necessary skills required for preparing and implementing research-based actions and interventions. Furthermore, they should develop a critical attitude towards the relevance and interpretation of (new) research data. On the other hand, social scientists should seriously strive to make domain-specific translations of new findings and insights for the professionals concerned and give answers to practical issues, in other words to conduct practice-based or practically-oriented research. A crucial question is what such practice-based research will look like and what the criteria will be that will have to be met to make research practice-relevant?

**The hegemony of internal validity and the Randomized Controlled Trial (RCT)**

Generally, evidence-based practice (EBP) is narrowed down to the use of results of so-called efficacy studies and cost-effectiveness research – both mainly quantitatively oriented – for the selection of interventions. In this context, the randomized controlled (or clinical) trial is generally considered to be gold standard of study design and, along with meta-analysis, ranks highest in Geddes (1999) hierarchy of - perhaps mainly efficacy - research. However, this type of research is only feasible under strictly controlled and well-defined settings and conditions, where the treatment or intervention to be tested is administered to an experimental group and will be compared with a control group. In addition, participants should be randomly assigned to the various groups (see Figure 1).
However, in the daily practice of social work and public health professionals, this type of research is hardly feasible - if not impossible. In a section aptly called ‘Overselling Randomization’, Howe (2004, pp. 45-46) seriously criticized both the use and the interpretation of the term ‘randomization’. In practical fields such as social work or public health, ‘random selection’ is often impossible; instead, what actually happens is a process of ‘random assignment’ of participants to experimental and control conditions. The resulting estimates, however unbiased, are thus restricted to a population of volunteers. Besides, due to political or legal considerations, ‘random assignment’ will often be impossible in the course of everyday practice.

Another valuable observation by Howe is his advice to trade external validity for internal validity in order to increase the generalizability of (the results of) practice-based research. The internal validity of a study pertains to the question whether a difference between groups can be attributed to the intervention or treatment or that it could also - partly - be attributed to external factors. Aiming to exclude these external factors, the researcher chooses to carefully define the group of respondents and to control the intervention or treatment as much as possible in order to optimize the study’s internal validity.

In this way, however, results are less likely to be externally valid; in other words, we would not know to what extent the results can be generalized to other persons, situations, conditions, measures, and measuring points than those which formed part of the specific study design. Moreover, as pointed out before, Howe’s critique of the feasibility of random selection and random assignment – two crucial cornerstones of internal validity – definitely does not strengthen the claims of the
supremacy of internal validity. In addition to Howe’s critical comments, one should be aware of the fact that the effectiveness of an intervention often cannot be proven unambiguously - i.e., by a simple ‘yes’ or ‘no’ (cf. De Los Reyes & Kazdin, 2008). I will come back to this later (see under Triangulation).

The American researcher and prevention expert Larry Green agrees with Howe’s critique when he argues that “[s]cientists are obsessively preoccupied to keep everything under control, in order to optimize the study’s internal validity. If they can make sure that a specified cause in their study has led to an effect they are satisfied, even if it would not be valid anywhere else” (cf. as cited in Van Megchelen, 2011, p. 9, translation and italics mine). Results obtained from this kind of artificial and contrived research - the ‘hothouse variety of evidence’ - will be only marginally useful for deriving guidelines for professional practices and activities, yet it is especially these studies that are being published on a large scale.

**Efficacy and effectiveness and ‘leakage’ in the scientific pipeline**

In a way, all of the afore-mentioned critical comments pertain to a crucial distinction: the difference between **efficacy** and **effectiveness**. Efficacy denotes the tested effect of an intervention under strictly controlled ‘laboratory’ conditions, whereas effectiveness denotes the tested effect of an intervention under more or less ‘normal’ conditions, as can be found in everyday professional practice, i.e., situations in real-time, with less control, and taking place in characteristic settings and contexts, populations and conditions (cf. Flay, 1986; Pittler & White, 1999; Green & Kreuter, 2005; Green & Glasgow, 2006).

Green et al. (2001, 2006a, 2006b, 2008; Green, Ottoson, Garcia & Hiatt, 2009) depict the process of knowledge transfer, dissemination of research results and the adoption and implementation of evidence-based guidelines in efficacy research as a pipeline, where scientific evidence will be produced that, after profound control and verification, will be distributed among policymakers, professionals and practitioners (see Figure 2).

A disquieting and frequently quoted statement about the overall leakage in the pipeline and the interval – or rather the loss – of time between (bio)medical research and medical practice means that it takes as many as 17 years to use only 14 % of the original research in actual practice (Balas & Boren, 2000)! Not only acquiring, funding and carrying out research takes quite some time, but a considerable amount of time will also get lost after completing the research, i.e., time needed for writing, submitting and revising articles, and waiting for publication. Subsequently, it will take time for results to be indexed in databases and systematic reviews, and to be recommended for ‘best practices’ in textbooks. Eventually, these practices will then
be implemented in public health institutions.

For public health professionals and practitioners in particular there are more specific disadvantages associated with the leakage in the pipeline. As a case in point, so-called ‘negative’ results are often not submitted by researchers, as they assume that they will not be published. Yet, for public health professionals and practitioners it could potentially be revealing to be well-informed about such ‘negative’ results, as they might yield insight into the (un)suitability of an intervention or method for certain populations and/or conditions – that may differ from the original (target) population and conditions.

Another important disadvantage for public health workers may occur by the loss of manuscripts that are rejected due to methodological ‘shortcomings’ regarding research design, statistical power or sample size. The largest losses, however, occur near the end of the pipeline: for (bio) medical research an estimated 6 to 13-year interval may pass between indexing research results and incorporating them in reviews that might lead to guidelines for ‘best practices’. This sizeable loss of time is due to the preponderant position of RCTs, which plays a decisive role in the final choice of results that will be included in systematic reviews, guidelines, textbooks, and other instructional materials for social workers and public health professionals.

Restraining influence of RCTs

In this way, the strictly controlled classic RCT works as a restraining factor, in that it keeps potentially relevant information about common professional practice in
everyday situations away from public health workers and officials. And, once again, the gap between research and practice manifests itself very clearly by the fact that - in (bio)medical research - it will, on average, take nine years (!) before evidence-based practices that found their way into reviews, directions for professionals, textbooks etc. will be completely implemented (see Figure 3).

![Figure 3: Gap between research and practice in (bio)medical research (Source: Green, 2010)](image)

Evidence that is tightened up, sieved out, highly controlled and tested in such a strict manner may work well for *biomedical interventions* that are often based on epidemiological research carried out with very large numbers of cases. However, in the field of primary and public (mental) health care and social work, the purpose of interventions is considerably more diverse as regards psychological processes, cultural context and socioeconomic conditions that may modify the relation between intervention and outcomes in many different ways. For these types of interventions, it is the context and setting, flexibility and external validity that are just as important as strict experimental control, precision of implementation and emphasis on internal validity are for efficacy research.

Practice-based research is strongly bound by methods that do justice to the complex reality of the work done by public health professionals and practitioners. In order to improve the quality of the work of these professionals, we need to study the effects of interventions in the context of everyday practice for which *effectiveness* studies are clearly the most relevant. In the next section I will, therefore, introduce and briefly describe several alternatives to RCT.
Alternatives to RCT: methodological remedies for practice-based research

In recent years, the appreciation and use of alternative research designs and data sources that will overcome the afore-mentioned disadvantages and limitations of RCTs has grown. Several years ago Van Yperen (2003) already had cast doubt on the pertinence of the RCT’s aura as gold standard of study design in the context of practice-based research. One of the alternative approaches he proposed is the multi-center study design (o.c., p. 37).

Multi-center study. In this study design, a number of institutions apply a new intervention, treatment or method, whereas other, similar, institutions with comparable clients or patients continue to apply the traditional method (‘care as usual’, CAU). Both client/patient groups will be matched in a number of relevant characteristics such as gender, age, type and severity of problems. In time, for instance after one or two years, the treatment groups will be compared with the CAU-groups from the ‘traditional’ institutions. A third group could be included, e.g., a wait list control group that did not (yet) receive any treatment. In this way, results of the new intervention can be compared with the CAU-group and an untreated group. This multi-center design allows for testing the effects of contextual variations between the participating institutions.

Observational studies. An observational study draws inferences about the possible effect of a treatment on subjects, where the assignment of subjects into a treated group versus a control group is outside the control of the investigator. This means that the ‘treatment’ that each subject receives is beyond the control of the investigator. A well-known example is the case-control study, originally developed in epidemiology, in which two existing groups differing in outcome are identified and compared on the basis of some supposed causal attribute, e.g., the effect of smoking on lung cancer.

Shrier et al. (2007) dispute the statement that RCT is a more valid study design for causal inference compared with the observational study design. They studied the effect of omitting and adding observational study results on RCT-results in meta-analyses and concluded that “(...) including information from observational studies may improve the inference based on only randomized trials”, and even claimed that “meta-analyses based on observational studies generally produce estimates of effect similar to those from meta-analyses based on randomized controlled trials”. In short, observational studies could be very valuable and the results they produce measure up to those from RCTs and should not be considered inferior or second-rate (cf. Rawlins, 2008).
Triangulation. Triangulation is the use of different methods or data sources in a study to cross-check results, for instance, comparing respondents’ interview data with their answers to items in a questionnaire, or think of a detective interviewing several different informants to examine a suspect’s doings and dealings.

Because the effectiveness of an intervention often cannot be proven unambiguously, i.e., by a simple ‘yes’ or ‘no’, De Los Reyes and Kazdin (2008) developed an elaborate and systematic triangulation procedure, called the Range of Possible Changes (RPC)-model. The model provides a classification system that identifies evidence-based treatments (EBTs) based in part on whether multiple or specific outcome methods consistently yield similar conclusions. The model subdivides EBTs into six categories and then cross-checks within each category results originating from different informants and obtained with different measuring instruments and methods. The highest category - denoting consistent evidence across multiple ways of gauging outcomes - is the Best Evidence for Change category. This category should be applied if “at least 80% of the findings from three or more informants, measures, and analytic methods show differences, and at least three findings were gleaned from each of the informants, measures, and methods. There is no clear informant-specific, measure-specific, or method-specific pattern of findings. The evidence suggests the intervention successfully targets the construct” (This category and the entire table can be found in De Los Reyes & Kazdin, o.c., p. 49).

Most critically, the RPC Model can be used to examine whether two studies of the same treatment yield consistent evidence between them. De Los Reyes and Kazdin (o.c., p. 48) give an example of two studies examining whether a particular treatment reduces symptoms of anxiety. If the studies could both be classified within the same category (e.g., Best Evidence for Change), then they may be classified as providing consistent evidence for the reduction of anxiety symptoms.

Mixed methods. The term mixed methods (or mixed research) denotes research that combines the use of both qualitative methods and quantitative methods and can be considered a form of methodological triangulation (Erzberger & Kelle, 2003, p. 458). The growing popularity of mixed methods in practice-based research stems from the assumption that a good mixed methods project will yield a more comprehensive understanding of the phenomenon under study and that it will produce more practice-relevant knowledge than could be gained from doing either qualitative or quantitative studies. The ‘real’, or purest, form of conducting mixed methods consists of an integration of qualitative and quantitative methods and results in all stages of the study. Approaches that are predominantly quantitative or qualitative in questions, and use two types of data without integration are actually quasi-mixed designs (Tashakkori & Teddlie, 2008).

A good example of the integrated mixed-methods design is the Poortinga et al. (2004) study of the food and mouth disease (FMD) outbreak in Great Britain. At the
height of the disease in 2001, the researchers conducted a survey by administering a 
self-completion questionnaire to two groups that were very differently affected by 
FMD. The questionnaire covered areas such as level of agreement with statements 
about the outbreak of FMD, perceptions of who was to blame, and degree of trust in 
various sources of information about the disease. In addition, a qualitative method 
(i.e., focus groups) was employed that covered the same areas as the questionnaire. 
Focus group participants were chosen from survey respondents who had indicated 
that they were willing to be involved in a focus group discussion. Three group 
discussions took place.

While the questionnaire data demonstrated the variation in such aspects as 
trust in various information sources, the focus groups revealed ‘valuable additional 
information, especially on the reasons, rationalizations and arguments behind 
people’s understanding of the FMD issue’ (Poortinga et al., o.c., p. 86). In this way, 
the researchers arrived at a more complete account of the FMD crisis than could 
have been obtained by taking either a quantitative or a qualitative research approach 
alone (Bryman, 2012, pp. 37, 638). A comprehensive textbook on mixed methods 
was written by Tashakkori and Teddlie (2003) and in 2007 the Journal of Mixed 
Methods Research was established.

**Pragmatic RCT.** RCTs can be classified as ‘explanatory or pragmatic’ 
(Schwartz & Lellouch, 2009). As has been extensively discussed above, *explanatory* 
RCTs test *efficacy* in a research setting with highly selected participants and under 
highly controlled conditions. In contrast, *pragmatic* RCTs test *effectiveness* in 
everyday practice with relatively unselected participants and under flexible 
conditions; in this way, pragmatic RCTs can inform *decisions* about practice, and 
they will possess higher external validity.

The choice between explanatory or pragmatic trials has to do with the purpose 
of the study: Is one to aim at an immediate increase in knowledge in the hope of 
eventual practical applications, or at a result which is of immediate applicability but 
which is less well understood and possibly less fertile for future (scientific knowledge) 
development? Or, in the words of Schwartz and Lellouch (o.c., p. 499), “[s]hould one 
prefer the goal of immediate applicability with a sacrifice of true understanding, or the 
more distant goal which may lead to greater enlightenment and which may prove 
more fertile for the future?” They also point out the importance of the *patient’s perspective* by underlining that “[a]gain, ethical considerations dominate - the type of 
trial must be chosen which is to the greatest benefit of the patients, both those in the 
trial and others” (Schwartz & Lellouch, ibidem).

A fine example of pragmatic trial methodology can be found in the Dutch 
Quattro Study, a pragmatic trial for evaluating the effectiveness of multidisciplinary 
patient care teams for the prevention of cardiovascular risk in primary health care 
practice in deprived neighborhoods in Rotterdam and The Hague, the Netherlands 
(Jansen, 2012).
Most trials conducted hitherto have adopted the explanatory approach without serious question or debate; the pragmatic trial approach would often have been more justifiable, not only from the patient's perspective, but also for the sake of obtaining externally valid results.

**Conclusion**

The objective of fundamental (basic) research into detecting building blocks for *evidence-based practice* can be characterized by the principal question whether the research outcome will be (1) internally valid, next whether it is (2) important and, finally, whether it will be (3) useful (cf. Figure 4).

From the perspective of practice-based research, a high price has to be paid, given that:

\[
\text{High internal validity} + \text{Low external validity} = \text{Diminished relevance for practice or policy}
\]

**Figure 4: Research in the context of Evidence-based Practice (Adapted from Ramirez, 2009)**

On the other hand, the objective of practice-based research into finding building blocks for *practice-based evidence* can be characterized by the principal question whether the research outcome will be (1) useful, next whether it is (2) important and, finally, whether it will be (3) sufficiently valid (cf. Figure 5).

Practice-based research will clearly benefit from this objective, given that:
High external validity + Relaxed internal validity = Increased relevance for practice

**Figure 5: Research in the context of Practice-based Evidence (Adapted from Ramirez, 2009)**

‘Increased relevance for practice’ has been a *leitmotiv* throughout this article, with the restriction and from the perspective that evidence-based practice should not be pushed aside or omitted but rather be supplemented and, if need be, replaced by research designs and methods that are closer to professional practice, yet retain sufficient evidential value.

It remains to be seen whether we really need the much debated term ‘paradigm shift’ (cf. Figure 5) to characterize the reduced emphasis on internal validity versus the increased importance of external validity in the context of practice-based evidence.

In this article I have brought the hegemony of internal validity and RCT up for discussion and have made a critical analysis on the subject. In (bio)medical research a serious debate on these matters has been going on for the last 15-20 years, but in many areas of (applied) social science research the ‘classic’ explanatory RCT is still considered the gold - and hitherto too much taken for granted - standard of study design.
References


Flay, B.R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. Preventive Medicine, 15, 451-474.


Green, L.W. (2010, November 4). Integrating Knowledge and Evidence in Public Health Policy & Practice: If we want more evidence-based practice, we need more practice-based evidence. Presented as a keynote address at the 3rd European Public Health Conference. Amsterdam: The Netherlands.


