

## External Validity in Research on Rehabilitative Interventions: Issues for Knowledge Translation

Marcel P. J. M. Dijkers, PhD

### Abstract

This issue of *FOCUS* discusses external validity and what rehabilitation<sup>1</sup> researchers can do to help practitioners answer the question “*How far can we generalize this finding—is it applicable to other clients/patients, with different characteristics, in dissimilar settings treated by other clinicians?*,” which clinicians and other practitioners ask whenever researchers publish evidence in support of a new or revised intervention. It is argued that there is no statistical basis for generalizing from the findings of a particular study, however well conducted, to a clinician’s next patient/client or even to the groups a health care policy maker is responsible for. Generalizing is done by the decision maker based on the similarity of the case or cases to be treated to the people who were included in the research, and based on the resemblance of the clinicians and rehabilitative settings involved to the treater(s) and setting(s) in the research.

Therefore, it is the responsibility of rehabilitation researchers to provide extensive information on their intervention study in their reports, well beyond issues impacting its internal validity. Only by reading such detailed reports on the characteristics of a study’s patients/clients, treaters, treatments and outcomes,

as well as on the wider health care and socio-political context of the study, can clinicians and their patients/clients begin to evaluate whether the same intervention might have a comparable feasibility and similar effect in their local situation. Unfortunately, reporting of issues relevant to external validity is much less common than of details relevant to internal validity, resulting in complaints by clinicians and other practitioners. This article provides and describes a checklist of 27 items that was designed to help rehabilitation researchers include more informative details in their research reports.

### Internal vs. external validity

It has been claimed that Campbell and Stanley (1963) originated the terms internal and external validity. In their influential monograph “*Experimental and quasi-experimental designs for research*” they defined *internal validity* as “the basic minimum without which any experiment is uninterpretable: Did in fact the experimental treatments make a difference in this specific experimental instance?” (p. 5). *External validity* was stated to concern generalizability: “To what populations, settings, treatment variables, and measurement variables can this effect be generalized?” (p. 5).

<sup>1</sup> When the term “rehabilitation” is used in this article, it refers to organized efforts to improve the functioning of people with mental and/or physical impairments and resulting activity limitations and participation restrictions. The framework used is within the context of medical rehabilitation but the concepts presented here may generalize to other areas of disability and rehabilitation research.

While internal and external validity are usually discussed in relation to randomized controlled trials (RCTs) and other controlled studies, they are equally relevant to other designs used to study the impact of treatments, such as pre-post studies and single subject designs. Whenever research on the impact of an intervention is reported, it is legitimate to ask two questions: 1. *How strong is the evidence that this particular intervention brought about these specific outcomes in the patients or clients treated?* and 2. *Even if there is solid proof that in this particular instance an improvement was a result of the (experimental) treatment, how far can we generalize this finding—is it applicable to other clients/patients, with different characteristics, in dissimilar settings, treated by other clinicians?* Specifically, it is appropriate for practitioners, the patients or clients they serve, and for policy makers and others who make decisions on treatments for individuals or groups, to ask the question: *Is this finding applicable to me, my patients/clients, the health services recipients I am serving?*

Unfortunately, it seems that in recent years the issue of internal validity has received a lot more attention than its twin, external validity. That is due largely to the turn the evidence-based practice

(EBP) movement has taken, especially within medicine and the health and rehabilitative sciences. In order to distinguish stronger from weaker evidence, EBP adherents sort research designs in a hierarchy, with the strongest on top and the weakest at the bottom. A representative hierarchy, that of the Oxford Centre for Evidence-based Medicine Levels of Evidence (2009), is presented in Table 1. It is easy to see that the single basis for classifying evidence is internal validity.

The same happens when systematic reviewers try to assess the quality of individual studies—whether those are RCTs or uncontrolled studies. For instance, the well-known PEDro scale consists of 11 items, 7 of which are focused on internal validity (Maher, Sherrington, Herbert, Moseley, & Elkins, 2003). There are four items that are relevant to external validity, but those seem to be there almost by accident, given the claim on the PEDro website “The PEDro scale considers two aspects of trial quality, namely the ‘believability’ (or ‘internal validity’) of the trial and whether the trial contains sufficient statistical information to make it interpretable. It does not rate the ‘meaningfulness’ (or ‘generalisability’ or ‘external validity’) of the trial, or the size of the treatment effect” (Centre for Evidence-Based Physiotherapy, 2011).

**Table 1**  
*A typical evidence-based practice (EBP) hierarchy of intervention research designs*

1a. Systematic review (with homogeneity) of randomized controlled trials
1b. Individual randomized controlled trials (with narrow confidence intervals)
1c. All or none survival
2a. Systematic review (with homogeneity) of cohort studies
2b. Individual cohort study (including low quality randomized controlled trial; e.g., <80% follow-up)
2c. "Outcomes" research; ecological studies
3a. Systematic review (with homogeneity) of case-control studies
3b. Individual case-control study
4. Case-series (and poor quality cohort and case-control studies)
5. Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"

From: Oxford Centre for Evidence-based Medicine. (2009). *Oxford Centre for Evidence-based Medicine - Levels of Evidence* (March 2009). Retrieved from <http://www.cebm.net/?o=1025>. Adapted by SEDL/NCDDR with permission from Oxford Centre for Evidence-based Medicine.

The emphasis on the internal validity of research on interventions goes further than grading the quality of studies. Some systematic reviewers are likely to consider anything weaker than an RCT to offer no evidence, and many of such reviews in *The Cochrane Library (The Cochrane Collaboration—Cochrane Reviews, 2010)* are “empty” reviews because the authors were unwilling to consider any empirical research that was not a (well-executed) RCT. Even systematic reviewers who do not throw out evidence resulting from, for example, a pretest-posttest study may consider this evidence less valuable, whatever the other good qualities of the research, e.g. its external validity.

This development offers problems for rehabilitation researchers, as they often cannot use the strongest designs in the hierarchies. By the nature of the interventions they study, practitioners (and often patients/clients) commonly cannot be blinded, random assignment may be unethical, etc. Even if researchers use all their skills to make the evidence as strong as it can be, given the circumstances, the EBP rating mechanisms and the ties between the grade of evidence and the strength of recommendations result in “weak” recommendations in almost all rehabilitative clinical practice guidelines and similar documents.

Not surprisingly, rehabilitation researchers desire to send their findings into the world with as strong a “good housekeeping” approval from EBP systematic reviewers as they can obtain, and they increasingly focus on internal validity in research design, research implementation (especially data collection) and research reporting—often under the banner of “rigor.” Frequently that means that everything that is relevant to “generalizability” gets short shrift. And that in turn creates problems for

practitioners (and their clients/patients, as well as administrators and policy makers): they may not get enough of the information they need to make an informed decision about adoption of the interventions the researchers claim to have proven to be effective.

Rehabilitative interventions typically are what have been termed “complex interventions” (Craig et al., 2008). They often involve multiple “essential ingredients” in addition to several other active ingredients (e.g. the practitioner-patient/client relationship), and entail numerous treatment sessions with carefully calibrated increases in

demands on the patient/client. These sessions themselves may be embedded in a package of rehabilitative treatments involving multiple disciplines that in turn are delivered as part of a complex arrangement of health and social services. They

have none of the simplicity of medications, which easily can be provided to clients/patients with a list of instructions on how to take them and what side effects to anticipate. But like psychotherapy and surgery, rehabilitation cannot be delivered that simply, and adopting a new treatment technique may require extensive training of the clinician and many changes in the delivery system to make sure that the presumed value of the new is actually realized. The implication is that rehabilitation researchers cannot simply name a new treatment or summarize it in a paragraph, and then spend five pages providing evidence that it brought about better outcomes (Dijkers et al., 2002). If they are serious about having practitioners use their findings, they must pay attention to issues of external validity. Which brings the issue back to Campbell and Stanley’s internal and external validity.

---

***Rehabilitation interventions typically are what have been termed “complex interventions” (Craig et al., 2008). They often involve multiple “essential ingredients” in addition to several other active ingredients (e.g. the therapist-patient relationship), and entail numerous treatment sessions with carefully calibrated increases in demands on the patient.***

## Clinicians' dissatisfaction with research reports

The EBP movement in health care and other fields has focused on the generation and synthesis of evidence and its well-considered use in clinical and other professional activities. As was indicated, in evaluating and synthesizing evidence, EBP scholars have concentrated on the internal validity of intervention research, to the almost complete exclusion of attention to external validity. The commonly known evidence hierarchies, e.g. of the American Academy of Neurology (Edlund, Gronseth, So, & Franklin, 2004), and the Oxford Centre for Evidence-based Medicine (2009), and the rankings produced by the most frequently used evidence quality rating instruments, e.g. the Jadad scale (Jadad et al., 1996) and the PEDro measure (Maher et al., 2003), are (almost) exclusively predicated on internal validity issues. Similarly, the various reporting guidelines for intervention research tend to give external validity short shrift. For instance, the original CONSORT (CONsolidation of Standards for Reporting Trials) statement contained only one element that mentioned generalizing from the study in question to wider populations (versus about nine addressing design elements relevant to internal validity). That one element was not very specific: "State specific interpretation of study findings, including sources of bias and imprecision (internal validity) and *discussion of external validity*, including appropriate quantitative measures when possible" (Begg et al., 1996; p. 638; italics added). It should be noted that much of the information called for by CONSORT, while not specifying external validity, is in fact very relevant to decision makers' ability to determine whether an intervention might apply to their situation. This will be discussed further in a later section. A wider interpretation of the concept of external validity,

---

***The RE-AIM framework is the only reporting guideline that truly focuses on describing those aspects of clinical research that are of direct relevance to external validity.***

as proposed in the current paper, would find that about eight prescriptions in the original CONSORT statement concern external validity.

The revised CONSORT statement, published in 2001 (Moher, Schulz, & Altman for the CONSORT Group (Consolidated Standards of Reporting Trials), 2001), did not offer much of an improvement with regard to external validity. With the introduction of supplemental CONSORT statements on the reporting of non-pharmacologic treatments (Boutron et al., 2008) and of pragmatic trials (Zwarenstein et al., 2008)

the situation has improved somewhat. For instance, the latter has this guidance: "Generalisability (external validity) of the trial findings: Describe key aspects of the setting which determined the trial results. Discuss possible differences in other settings where clinical traditions, health

service organisation, staffing, or resources may vary from those of the trial" (Zwarenstein et al., 2008; p. 3).

The TREND statement (Des Jarlais, Lyles, Crepaz, & the TREND Group, 2004) and the SQUIRE guidelines (Davidoff et al., 2009) are somewhat better than the (original) CONSORT. The RE-AIM framework (Glasgow, Vogt, & Boles, 1999), known primarily in behavioral medicine circles, is the only reporting guideline that truly focuses on describing those aspects of clinical research that are of direct relevance to external validity.

If researchers focus the reports of their studies on those aspects of research design and implementation that affect internal validity, encouraged in doing so by semi-official statements aiming to improve reporting of interventional research, it is not surprising that clinicians and other practitioners complain that the published literature is not very useful to them. It is worth repeating a line from Rothwell: "Lack

of consideration of external validity is the most frequent criticism by clinicians of RCTs, systematic reviews, and guidelines” (2005, p. 82). There are two basic reasons for this criticism.

One is related to the *design and nature* of the studies that are funded and that do get published. They primarily concern treatment situations that are *unlike* the ones clinicians deal with, due to commonly described factors such as: sample restrictions regarding age, sex, race, co-morbidities, etc. (Van Spall, Toren, Kiss, & Fowler, 2007); extensive use of methods to ensure clients'/patients' adherence to treatment which in reality no practitioner can apply; absence of functionally significant outcomes; disregard of patient/client preferences for or against specific treatment alternatives; and absence of assessments of cost and cost-effectiveness (Depp & Lebowitz, 2007). In terms of the continuum between *explanatory* and *pragmatic* studies, most published research is much closer to the first pole (Vallve, 2003).

Explanatory studies (sometimes called controlled laboratory, fastidious or regulatory studies) are designed to answer the efficacy question: “*Can it work?*” and as such focus on internal validity: demonstrating the impact of the treatment on the specified outcomes under “ideal world circumstances.” Pragmatic intervention studies (also called management, practical, large sample, health systems and public health model studies) are designed to answer the effectiveness questions: “*Will it work, and how well? Is it worth it?*” They focus equally on external validity, generalizability, applicability, transferability and extrapolation (Dekkers, von Elm, Algra, Romijn, & Vandenbroucke, 2010): demonstrating the potential of an intervention under “real-world” or usual circumstances. Explanatory and pragmatic

studies may differ from one another on a number of dimensions relevant to generalization, as described by Gartlehner, Hansen, Nissman, Lohr, & Carey (2006) and Thorpe et al. (2009).

A second reason for practitioners' dissatisfaction with treatment research is the *reporting* of this research. The focus in most intervention research papers is on the components that contributed to and reflect internal validity, but any information that might be of use to clinicians in assessing applicability of the intervention to their own

patients or clients gets limited space. This claim is backed up by some systematic review evidence, for instance the paper by Klesges, Dzewaltowski, and Glasgow (2008).

This type of complaint is infrequently found in the rehabilitation literature. It may be that rehabilitation

researchers do a better job of designing and reporting research that is useful to clinicians and other decision makers, or that rehabilitation practitioners are less prone to complain about the utility of the published research. More likely is that they hardly read it (Burke, DeVito, Schneider, Julien, & Judelson, 2004; Burke, Judelson, Schneider, DeVito, & Latta, 2002; Jette et al., 2003; Turner & Whitfield, 1997), and instead get their ideas for new interventions from colleagues and in seminars, rather than directly from the research literature (Philibert, Snyder, Judd, & Windsor, 2003; Powell & Case-Smith, 2003).

### **External validity as the impossible dream**

All treatment studies by necessity take place at a particular place and time, and involve a sample of all individuals who might benefit from the particular intervention being studied. In the best of all worlds researchers would create a list (a “sampling frame”) of all patients/clients worldwide who have the potential to benefit from the



treatment of interest, and draw and study a random sample. If the research demonstrated a robust effect of the treatment, this could be generalized to all individuals on the list. Unfortunately, the worldwide sampling frame is a theoretical concept only; in practice, researchers deal with a *convenience sample* of clients/patients available at a particular place and time, and as a consequence there is no *statistical* basis for generalizing to patients/clients elsewhere and at other (future) times, who may differ from those studied in terms of physiology, culture, severity and presentation of the problem of interest, and in many other ways that may affect the adherence to and effectiveness of the treatment being studied.

It is perhaps little appreciated that the *same* constraints on generalizing hold true for all other aspects of the study. While a medication may have the same effect on people worldwide, and can be reproduced in brand-name and generic copies to be used wherever needed, the same is not true for all other interventions. Surgery, behavioral medicine, psychotherapy and most forms of rehabilitation involve a treater who develops a relationship with a client/patient in a particular setting and delivers a complex treatment, often involving multiple sessions and auxiliary services

delivered by an entire team. Any one study can only assess the effectiveness of a particular selection of the multiple, potentially equally valid, constellations that can be created by selecting a few treaters in one or two settings who dutifully follow one particular protocol. Just as we cannot generalize from one patient/client sample of convenience to all patients/clients, so we cannot generalize from one treater sample (sometimes an N=1 sample), one setting sample, one protocol sample, to the bewildering variety of settings in which treaters with a range of disciplinary backgrounds and levels of expertise may deliver slight variations of the same treatment.

Similar restrictions hold true with respect to the third major component in study design: the outcomes. While researchers may talk about the effect of a treatment on various domains of functioning or quality of life, in order to *demonstrate* the effectiveness of the intervention they have to select a particular outcome measure, which by necessity is limited to a few indicators of the outcome domain. While in theory the indicators that have been combined into a measure are a random sample of all possible indicators of a construct of interest, in practice they often constitute a sample of convenience, and there may

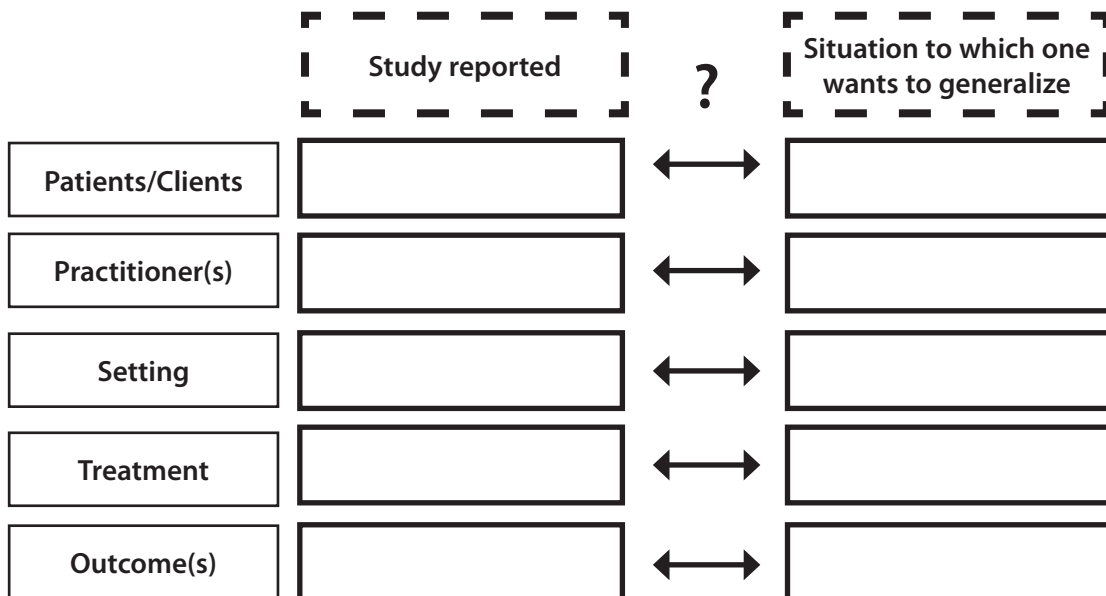
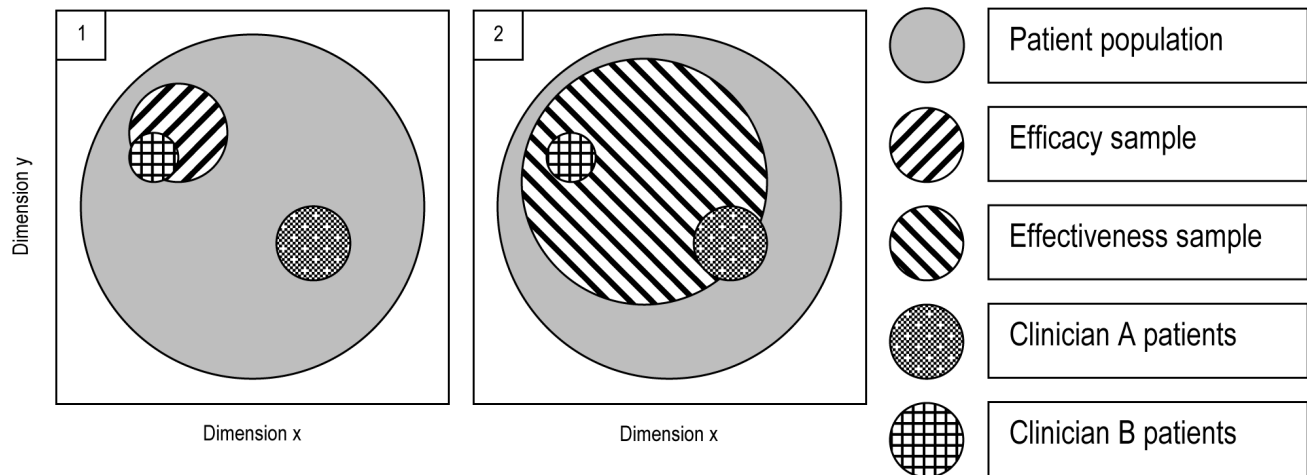


Figure 1. The basic dilemma of generalizing from intervention research to practical situations.



**Figure 2. Overlap between hypothetical efficacy and effectiveness studies' patient/client samples, and the clientele of clinicians A and B**

be a limited basis for generalizing to the entirety of the domain of interest, let alone to other domains.

The fact that there is no *statistical* basis for generalizing from one particular treatment study's results to all clients/patients, all settings, all treaters and all similar outcomes, puts a special burden on decision makers. They have to examine the report of the study and answer a very difficult question: *"Are the subjects, settings, treaters, procedures and outcomes of this research similar enough to my patients (clients), my site(s), my practitioners, my feasible procedures and the treatment effects that are of interest to me and to my patients/clients, that it is reasonable to assume that the intervention studied will have the same or very similar effects in my situation?"* (see Figure 1).

Answering the question requires a great deal of information about the study in question, and on one's own situation, as well as a great deal of clinical expertise built on a solid basis of didactic and experiential knowledge. Researchers can contribute only one part of what is needed to make the decision: information on the study. Based on what they know to be or assume to be the relevant information needs of the *potential* users of their findings (decision makers of one or many stripes) they have to provide sufficient details on *who* was treated with *what intervention* in *what settings* by *what clinicians* to reach *what positive outcomes*.

**The utility of reports of explanatory vs. pragmatic trials for decision makers**

Sometimes the claim is made that the results of explanatory trials are of no or at best very limited use to clinicians, but that those of pragmatic trials are "by definition" of great utility. However, that is an erroneous conclusion from a comparison of the two types of studies; the fact that there is a *continuum* of studies that is based on differences in *all* the dimensions listed by Gartlehner et al. (2006) and Thorpe et al. (2009) should make that clear. At best one can put forward the claim that pragmatic studies are *more likely* to be of benefit, to more categories of decision makers. Figure 2 may clarify this; it only addresses the match between the research situation and the setting one wants to generalize to in terms of nature of the clients/patients. In order for a study to be informative to decision making for specific individuals and groups, there has to be a match between the subjects in the study and the patients/clients of the decision maker, in terms of the various characteristics that are assumed to be relevant to the treatment. Because in a pragmatic trial typically a larger variety of patients/clients is included, there is a better correspondence with the entire heterogeneous client or patient population that is thought to have benefit from the treatment (the solid grey circles in the two panels of Figure 2), and a greater likelihood that the types of patients/

clients one needs to make decisions for were included in the study.

The patients/clients of Clinician A in panel 2 in the Figure have an overlap with the subjects studied in the effectiveness trial, as have those of Clinician B, who tend to be quite different in their relevant characteristics. However, while in the explanatory study depicted in panel 1 none of the type of clients/patients Clinician A has are included, Clinician B's clients/patients are represented. If there is similar overlap or exact match in other relevant aspects (practitioner, treatment setting and delivery, and outcomes—see Figure 1) between the panel 1 study and the setting and treatment characteristics as well as important outcomes of Clinician B, he or she may be able to use the findings of the “narrow” explanatory study. Thus, *the perspective of the potential user* is all-important. If there is a report on an explanatory study, however “narrow”, that happens to match the patients or clients treatments and outcomes that the user is considering, then she or he can use the information in making decisions. On the other hand, if a pragmatic study—however “wide”—fails to match the practitioner's needs, whether it is in terms of patient/client characteristics, interventions the practitioner is qualified to deliver, outcomes or any other issue, then the study findings are of no use to that particular reader.

A corollary of this conclusion is that in order for the practitioner, patient/client, or policy maker to be able to make a decision, they need to have extensive information on *all* aspects of the study in question, and it is the researcher's responsibility to provide it. “We would argue, therefore, that trialists should not worry about trying to guess the various perspectives of those making decisions but should instead do all they can to describe the context of their trial” (Treweek & Zwarenstein, 2009, p. 5).

With context, Treweek and Zwarenstein mean “the distinctive features of a trial's setting, participants, clinicians and other staff” (p. 5). In essence, it does not matter whether one's study is an explanatory study or a pragmatic one, or a careful blend of aspects of both. In order for clinicians to be able to

---

***A corollary of this conclusion is that in order for the therapist, patient, or policy maker to be able to make a decision, they need to have extensive information on all aspects of the study in question—and it is the researcher's responsibility to provide it.***

use the results in their treatment of all or some of their patients or clients, researchers must collect the necessary information, and put it in their journal report or make it available in supplemental materials. Without it, the first step toward knowledge translation is a misstep. “An internally valid trial that has poor applicability, or is reported in such a way that it is difficult or

impossible for others to make judgements about its applicability, is a lost opportunity to influence clinical practice and healthcare delivery” state Treweek and Zwarenstein (2009, p. 7).

Gartlehner and colleagues (2009) summarize the situation as follows: there are three questions clinicians need to have answered before they can apply a new or modified treatment to their patients:

- Is the study population similar to my patient population?
- Does the study intervention resemble my clinical practice, or my practice with minor modifications?
- Are the outcomes reported relevant to my decision making with or for my patients—benefits, harms and costs?

Kendrick, Hegarty, and Glasziou (2008) offer a longer list, focused on the individual patient rather than a caseload. Only clinicians can answer those types of questions with yes or no, and they need two matching sets of information. It is the researchers' responsibility to provide half the information; the other half (information about their local patients/clients and practices) is supplied by the practitioners themselves (see Figure 1). The



researcher needs to be specific; including in one's research report such phrases as "typical patients received usual care" is not going to be of much use—in the end, there may be major differences between the patients/clients, practices, health care systems and cultures of the various practitioners who read the report, and "usual" means something different to all of them.

This leads the researcher to the question: *then what exactly should I report, and in what level of detail?* It is unreasonable to expect investigators to report or busy clinicians to read every last detail that might ever be desired by any potential user of research findings; research reports would run 50,000 words instead of the 5,000 now allowed by the more generous journal editors.

### **A checklist of external validity items for reporting explanatory and pragmatic research**

Methodologists and clinical researchers who have considered external validity issues have begun to answer the questions related to what information is needed to support transfer of one's research findings to decision makers. The checklist in Table 2 is based on the suggestions of a number of these scholars (Ahmad et al., 2009; Bonell, Oakley, Hargreaves, Strange, & Rees, 2006; Bornhoft et al., 2006; Boutron et al., 2008; Davidson et al., 2003; Dekkers et al., 2010; Des Jarlais et al., 2004; Dzewaltowski, Estabrooks, Klesges, Bull, & Glasgow, 2004; Glasgow, Bull, Gillette, Klesges, & Dzewaltowski, 2002; Glasgow, Davidson, Dobkin, Ockene, & Spring, 2006; Glasgow, Klesges, Dzewaltowski, Estabrooks, & Vogt, 2006; Glasziou, Meats, Heneghan, & Shepperd, 2008; Jacquier, Boutron, Moher, Roy, & Ravaud, 2006; Klesges, Estabrooks, Dzewaltowski, Bull, & Glasgow, 2005; Leeman, Jackson, & Sandelowski, 2006; Ogrinc et al., 2008; Rothwell, 2005; Tansella, Thornicroft, Barbui, Cipriani, & Saraceno, 2006; Zwarenstein et al., 2008).

This list is a lengthy one, and detailing information relevant to each item in one's report might seem to result in papers long enough to guarantee

rejection by all journal editors. However, it should be kept in mind that "rehabilitation research" involves a wide scope of intervention types and targets—from individual muscles in some physical therapy treatments to attitudes of groups of patients in rehabilitation psychology. What is important to report in one type of study may be inconsequential or not applicable in another. A behavioral treatment as might be used by a rehabilitation psychologist needs an exquisite description of the protocol as well as the training for the practitioners involved. On the other hand, a medication treatment for spasticity provided by a physician can be described by providing a drug name and dosage schedule; the expertise of the person handing out the pills is hardly important, unless careful monitoring for side effects is needed as part of titration. The checklist aims to cover all these situations, and every entry should be considered to be preceded by the phrase: "if any/ if applicable." Secondly, these are not all new items. A number of them already are routinely reported in much of rehabilitation research, although the authors may never have given a great deal of thought to the importance of reporting on issues especially relevant to external validity. For many of these items, some additional detail may be the most that is needed to make one's research report more useful.

It should be noted that the list is not just applicable to the reporting of controlled studies. If a study with a weaker design, a single-group pre- vs. post-intervention comparison for instance, is worth reporting, it is worth reporting in the detail that allows for well-founded adoption of the intervention by others. Uncontrolled designs have weak internal validity according to methodology textbooks, but in rehabilitation there may occur situations that are the exception to the rule. Providing a wheelchair or a prosthesis may make such an instantaneous and dramatic difference in patients' functioning that a control group is not necessary to demonstrate effect (Johnston, Sherer, & Whyte, 2006); it parallels category 1c evidence (all or none survival) in Table 1.

**Table 2**

*Checklist of items rehabilitation researchers should include in their manuscripts to achieve maximal adoption potential*

CATEGORY	ITEMS POTENTIALLY REQUIRING DESCRIPTION	✓
<b>1. Rationale</b>	Health/functional problem or health/rehabilitative services problem addressed by the intervention; Other interventions commonly used for this problem (whether or not comparator in the study)	
<b>2. Target population</b>	Characteristics of all clients/patients for whom the intervention might be/is assumed to be applicable: clinical, cultural and sociodemographic characteristics	
<b>3. Recruitment of clients/patients</b>	Method (referral sources; advertising, etc.); Duration of recruitment period; Patient/client incentives offered	
<b>4. Patient/client inclusion criteria</b>	Type, stage, duration and severity of target disorder/problem: diagnostic tests and assessments used; Functioning level, in various relevant domains; Justification of choices	
<b>5. Patient/client exclusion criteria</b>	Type and severity of comorbidities: diagnostic tests and assessments used; Current and prior treatments for the condition of interest and for comorbidities; Age and other demographic factors; Patient/client skills and abilities (e.g. literacy level): diagnostic tests and assessments used; Likelihood of compliance, non-response and adverse events (e.g. as determined by use of run-in periods); Current/recent participation in another trial; Justification of choices	
<b>6. Number of clients/patients</b>	Sample size: clinical minimally important difference or other clear and important clinical outcome as basis for power analysis; Total number screened; Number excluded per inclusion criteria; Number excluded per each/combined exclusion criteria; Number refusing informed consent; Number studied, by center and/or by treater; Number dropping out, by phase (treatment and follow-up); Patient/client flow diagram	
<b>7. Nature of clients/patients</b>	Clinical (e.g. severity and duration of disorder, co-morbidities, functioning level, medications) and demographic (e.g. age, sex, race/ethnicity, education, socio-economic status, marital status) characteristics of clients/patients included/randomized; Clinical and demographic characteristics of clients/patients excluded/of clients/patients refusing consent; Baseline risks of good/poor outcomes in the treatment/control group; Representativeness of patients of target population; Clients'/patients' therapy preferences and expectations	
<b>8. Substitute targets (e.g. parent for child)</b>	Relationship of patient/client to target of intervention; <b>All information on recruitment, inclusion/exclusion criteria, numbers, nature as listed for clients/patients (#3 - #7)</b>	

CATEGORY	ITEMS POTENTIALLY REQUIRING DESCRIPTION	✓
9. Nature and number of treating centers	Eligibility criteria for centers; Type and number of potential sites included/excluded; Primary, secondary or tertiary level of care; Number of clinicians and of clients/patients treated per year, overall; Number of clients/patients with the condition studied or closely related conditions treated per year	
10. Nature and number of treaters	Eligibility criteria for treaters; Years of experience overall and with the condition studied; Number of clients/patients with condition studied (and/or closely related conditions) treated in last few years; Number of clients/patients (in each study arm) by center/treater; Blinding of treaters	
11. Treater training and support	Nature of treater training in administering the treatment studied (type of training sessions, trainer, practice cases, etc.); Methods used to improve protocol adherence (supervision, meetings of treaters, incentives, etc.)	
12. Treatment and treatment delivery	Prerandomization diagnosis and assessment methods; Non-trial treatments allowed and disallowed; Whether those administering co-interventions were blinded to patient's study group assignment; Number, timing and duration of each session; total treatment time; Content of treatment components; availability of/nature and level of detail of written protocol; Brand and model of all specialized equipment used; Nature of all media provided (written instructions, videos, interactive computer program, etc.) to treaters and/or patients/clients; Medications: generic and proprietary name, titration and dosage, route and timing, duration, clinical monitoring for effects/side effects; Location(s) and settings of service delivery; Individual vs. group delivery; Local adaptations/modifications of the protocol used at participating sites; Treatment acceptability (other than refusal and drop-out): cultural and financial issues, patient/client preferences; Nature and duration of all other services/treatments/required Algorithms for intervention tailoring; Individualization to patient/client needs allowed; Methods for standardization used; Incentives and other mechanism to increase compliance	
13. Comparator	<b>Same items as described for treatment (#12)</b> Similarity of intervention and comparator from layman's perspective (if blinding was not possible)	
14. Wider context	Country Healthcare system Insurance companies' and regulatory authorities' policies limiting access to treatment or comparator, to needed diagnostic workups Years of research data collection New diagnostic and treatment methods developed since trial/study start	

Table 2, continued

Checklist of items rehabilitation researchers should include in their manuscripts to achieve maximal adoption

CATEGORY	ITEMS POTENTIALLY REQUIRING DESCRIPTION	✓
15. Treater compliance	Methods and results of measurement of treatment integrity: treater-completed checklists, assessments based on audio/videotapes; data from mechanical counters, etc.; Treater dropout; Treatment and comparator contamination	
16. Patient/client compliance	Methods and results of measurement of treatment integrity: patient/client-completed checklist/diaries; data from mechanical counters, etc.; Treatment discontinuation rates and reasons for dropout (adverse effects, treatment ineffectiveness, other)	
17. Outcome measures	Primary and secondary measures: biological, functioning, quality of life, provider and patient/client global change ratings, etc.; Instruments used and appropriateness for patient/client sample; Adverse events, negative outcomes, unintended consequences; Costs/economic outcomes (to clients/patients, systems, others); Nature, independence and blinding of outcome assessors	
18. Follow-up and outcomes	Duration of follow-up and assessment points; Outcomes at all time points	
19. Data analysis	Comparisons between participants and non-participants; Comparisons between participating and non-participating sites/treaters; Exclusions from exact intent-to-treat design, for each analysis; Imputation methods used; Outcomes for setting/treater/patient-client subgroups; Degree of contamination between treatment and comparator; Impact of non-adherence on outcomes; Blinding success; Sensitivity analyses for alternative assumptions	
20. Other issues	Availability of protocols and other treatment materials Continued use of the treatment at the sites/by the treaters; Reasons for post-study modification or discontinuation; Implementation theory and hypotheses; Implementation process, extent, development; Comparison of processes, if two active treatments	
21. Discussion	Clinical significance of results Treatment feasibility: contexts, professional and other workload, costs, equipment, treater training levels, treatment duration, treater/patient-client adherence, etc.; Choice of comparator, especially similarity/difference of therapeutic effect of patient/client relationship with treater; Acceptability of treatment: fit with current practice, fit with client/patient culture; etc.; Groups (clinical, demographic) for whom the treatment may be particularly suitable/unsuitable; Blinding failure, risk of bias related to non-blinding; Co-interventions, contamination impacts; Bias related to differential expertise of treaters in each study group	

Studies that are weak in terms of internal validity often compensate for that in terms of strong potential for external validity—the typical pragmatic vs. explanatory RCT, for instance. Whenever the level of internal validity of the study is acceptable to the researcher and the reader, information relevant to generalization should be provided to assist the latter in judging to whom and in what situations the intervention described may apply.

### **A short description/justification of the checklist items**

The first element on the list, the problem addressed (#1) by the experimental and comparator treatments (if any), is generally well covered in contemporary reports of research. Many authors are less specific in describing the exact target population for the intervention trialed, though, moving directly to a listing of the sample selection criteria. However, as the target population (#2) for the intervention may or may not coincide in terms of relevant characteristics with the group actually studied (see Figure 2), some explanation may be in order.

The methods of recruitment of subjects (#3) may appear irrelevant to potential users of the treatment studied, but they inform on the nature of these subjects. Mention of the duration of the recruitment period in combination with the numbers of clients/patients screened and enrolled (#6) similarly may provide information on how select the group is, although to make a judgment on that, the reader also will need information on the number of clients/patients with the disorder or problem of interest seen yearly by the centers and practitioners involved in the study (#9, #10).

The inclusion and exclusion criteria for the subjects enrolled in the research (#4, #5) are a familiar element of all clinical research reports, but research has shown their coverage may be incomplete (Shapiro, Weijer, & Freedman, 2000). Clearly, reports aimed at clinicians that omit information on co-morbidities and safety factors

run the risk of making routine care worse rather than better.

The surprising entry in the inclusion and exclusion criteria section (#4, #5) may be the “justification of choices” element. It is of little use to readers to know that an exclusion was justifiable; whether the exclusion destroys the comparability between the clients/patients studied and their own target group is the crucial issue. However, this one item is a reminder to investigators that in planning research they must have good reasons for each exclusion and inclusion criterion so as to put the proper emphasis on the explanatory vs. the pragmatic end of the efficacy-effectiveness continuum (Gartlehner et al., 2006; Thorpe et al., 2009). In many instances narrow inclusion criteria result in slow enrollment. Then, in order to achieve their targeted sample size, researchers frequently proceed to relax the criteria. If this makes sense in light of the study’s rationale (#1) and target population (#2), it would have made sense at the start of the project.

The exact number of subjects that made it to various stages of the research (recruiting and consenting; treatment; follow-up assessment of outcomes) (#6) is probably of key interest to readers. Specifically, they need to know not just the numbers of (potential) subjects but the nature of the different groups (#7): how many were excluded by various rule-in and rule-out criteria, how many by the requirement to give informed consent, and how did these groups compare? Given the fact that trials may exclude over 90% of all patients with the condition of interest, for various administrative, medical and scientific reasons (Britton et al., 1999), this information is crucial for decision makers in assessing the potential benefit of the treatment to their clientele. They need information on two questions especially: *How did those who satisfied all the criteria but did not consent differ from the ones who did? In the latter group, how did people who dropped out of the study compare with those who completed the final assessment?* In a pragmatic study even if intent-to-treat (ITT) analysis is used to arrive at a bottom-line conclusion, practitioners



may still have a strong interest in the clients/patients who dropped out, especially those whose discontinuation was related to the treatment—lack of effectiveness, unbearable side effects, etc.

“Representativeness of the target population” strictly speaking requires exact knowledge of the make-up of the target population and the study sample in terms of all clinical and other characteristics known or assumed to be relevant to the treatment or its outcomes. As was argued previously, this is not possible. However, authors may mention (with or without exact data to back up their observations) that particular subgroups appear to be missing or are over-represented, which may be important to know for some readers. In rehabilitation it is not uncommon that practitioners deal not just with clients/patients but also with their family or personal care assistant. If the study’s goal is to improve functioning of clients/patients through instruction or training of caregivers, these direct targets need description, as well as the ultimate beneficiaries of the intervention (#8).

Methodologists working in the areas of behavioral medicine and public health in recent years have emphasized the need to describe in more detail the nature of the treating centers (#9). Similar concerns are increasingly uttered in surgery research (Pibouleau, Boutron, Reeves, Nizard, & Ravaud, 2009). The number of procedures that surgical teams perform each year has been shown to affect outcomes, and there is no reason why the same would not hold true for rehabilitation. Certainly, many families of patients with spinal cord injury (SCI) or stroke, for example, understand this intuitively, and have on their list of criteria for choosing a rehabilitation facility the annual SCI or stroke patient volume of the centers they are considering.

Along similar lines, psychotherapy and surgery researchers have been paying attention to the characteristics of the individual clinicians in their studies (#10). Both their annual patient volume and number of years of experience may be of relevance. The training of the study clinicians in the treatment methods required by the study (if any), and the mechanisms the investigators put into place to make sure that they follow the protocol (if any) need to be described (#11) in order to inform readers of the resources that were needed to achieve the reported level of protocol adherence (#15).

In treatment research the intervention is as important as the outcomes, but rehabilitation researchers traditionally have paid relatively little attention to describing the former (Dijkers et al., 2002). While it may be unrealistic to expect that a well-experienced clinician can replicate a treatment based on a description in a research article (Marks, 2009, 2010), the basics for understanding the treatment should be there (#12). Additional information, and possibly training, should be available from the authors.

The prototypical treatment information is a listing of drug name, dosage and timing, but most rehabilitative treatments are much more involved, and as “complex interventions” (Craig et al., 2008), mix active ingredients of various classes and multiple sessions, possibly with various categories of clinicians involved. The schedule for titration of the active ingredient(s) and/or the content of successive sessions may need to be specified in rehabilitation research, as well as the location of the treatment (outpatient clinic, community health center, etc.) and the equipment used. While inpatient rehabilitation clinicians may have a monopoly in treating their patients (except for what their colleague of another discipline may provide the patient in the next

---

***Authors may mention (with or without exact data to back up their observations) that particular subgroups appear to be missing or are over-represented, which may be important to know for some readers.***

hour), those treating outpatients must be aware of the supplemental treatments patients may seek (complementary and alternative medicine, for instance), and may even administer or arrange for it, to get patients to their goals quicker. Such non-trial treatments may be allowed in a protocol; their nature and frequency certainly ought to be described. The treatments that the clients/patients enrolled in a study receive for problems and disorders other than the target of the study similarly need to be reported, because they may have synergistic or antagonistic effects.

In multi-site studies, local modifications of the protocol may be allowed to make a treatment a better fit with the culture of patients or with the patient/client flow patterns in a health care setting. The acceptability of treatments and its effect on adherence may need to be reported on a by-site basis, as may a number of other approved treater-to-treater or site-to-site differences, not just for the experimental treatment but also for the comparator (#13). It always should be assumed that a comparator has effects (even if nothing more than placebo or nocebo effects), and therefore a comparator should be described in as much detail as the treatment being studied. The choice of comparator has major implications for the findings of efficacy and effectiveness of the experimental intervention, and decision makers require all relevant detail (Hart, Fann, & Novack, 2008; Mohr et al., 2009).

While blinding of clients/patients to rehabilitative interventions often is impossible, researchers would seem to overestimate clients'/patients' ability to tell one highly involved therapy from another (for instance, Bobath interventions vs. Feldenkrais after stroke), and in effect blinding may be achieved through clients'/patients' lack of expertise. Certainly it is possible to ask clients/patients upon study completion what treatment they *think* they have received, and to describe (lack of) similarity of treatment and comparator in laymen's eyes. This is not just relevant to the impact of blinding on internal validity, but also to the adherence

to treatment, which is a key item of interest to potential clinician adaptors of a new treatment.

Rehabilitation research increasingly is becoming international, with non-US researchers publishing in American journals, and European English-language journals such as *Clinical Rehabilitation* and *Journal of Rehabilitation Medicine* easily available to US practitioners. However, it should be realized that rehabilitative interventions, especially those addressing participation restrictions rather than impairments (to use ICF language), depend for their success to a large degree on social and cultural factors, which do not cross borders as easily as the journals themselves. Health care systems and health insurance schemes may affect access to specific types of care. Cultural, economic and logistical factors may dictate the translation/transition potential of a newly proven intervention (Bagshaw & Bellomo, 2008), and researchers should make sure to describe the wider context of the intervention they are testing (#14).

Whether compliance by practitioners and clients/patients is strictly controlled (in explanatory studies) or left open to be measured as a study outcome (in pragmatic studies), the level of adherence is typically reported in both (#15, 16). More extensive tracking of protocol compliance is typical in explanatory studies. However, even effectiveness studies ought to report on the degree of compliance that was achieved, if only to give potential adaptors of the treatment a basis for estimating the level of treatment success they might have if more (or less) incentives for compliance are put into place. One systematic review of rehabilitation intervention studies completed a decade ago found minimal efforts by researchers to assess procedural reliability (Dijkers et al., 2002); a recent review of a narrower group of studies (Robb & Carpenter, 2009) did not offer evidence to indicate that the situation has improved.

Rehabilitation researchers have spent much effort (too much, some would say) to develop outcome

measures (#17), corresponding to each one of the impairments, activity limitations and participation restrictions outlined in the ICF, as well as outcomes in other domains. Given the nature of rehabilitation and rehabilitation research, the vast majority of these have been in the areas of life that clients/patients, clinicians and other decision makers consider important. Adverse events are not often reported, although the claim can be made that most rehabilitative interventions are such that they have no dramatic effects, but there is also not much of a chance of serious adverse effects. A major shortcoming (in rehabilitation research as well as scholarly work in other health care areas) is the lack of reporting of the costs of care, whether that is in terms of resource consumption per patient/client per episode of care or a more traditional economic calculation in terms of health care dollars that would need to be expended by future clients/patients or their third party payers. Outcome measures of *any* type may need to be reported for not just the end of active treatment, but also later on, to provide decision-makers with the treatment's potential for sustained effect (#18).

The report on the analysis (#19) of an intervention study's data involves more than specifying whether the statistical testing of the key outcomes is ITT or per-protocol. While the potential users need that information, they are also served by additional analyses of which factors may affect the outcomes (for instance, failure of blinding, attrition), and what the findings are for subgroups of clients/patients. Reports of the latter should be clearly labeled either as prespecified in the protocol, or as ad-hoc and suggested by the accumulated data.

A number of other issues (#20) may be relevant to a treatment's potential for adoption in routine practice. Some behavioral medicine scholars have stressed the importance of researchers reporting on the continued use of experimental treatments at the sites and by the clinicians involved in the study. If these individuals or facilities discontinued the intervention after the study

ended, that may be a clear indication that, even if the treatment was declared a success, it is not very viable. This could be due to high resource costs, or because of a poor cost vs. outcomes balance, or as a consequence of more—or more serious—side effects than patients or clinicians are willing to accept.

There are many aspects of any treatment that may be hard to capture in exact numbers, but that decision makers may want communicated just the same, as these may constitute the information that pushes a decision one way or another. When the researcher has no hard data, the Discussion section (#21) may be the appropriate place to bring up issues such as clinical significance in light of the study's effect size (Faulkner, Fidler, & Cumming, 2008); the fit of a newly proposed intervention to the culture of a discipline or a type of facility; and the authors' or the participating clinicians' judgment as to the subject groups with whom the treatment may be unsuccessful because of clinical, cultural or other factors.

### **Using the list in planning and reporting rehabilitation research**

Rehabilitation researchers considering use of this checklist should keep a number of things in mind. First of all, the entire list is applicable to all types of intervention research—controlled and uncontrolled, explanatory and pragmatic. As is illustrated in Figure 1, "external validity" is not an inherent characteristic of one's research findings, the way internal validity is. In the absence of random sampling of clients/patients, treaters, protocols, settings and outcomes, it is the needs and the situation of a potential user of research findings that determine applicability to a practice setting. The researcher should provide the practitioner with necessary information about clients/patients and setting, treatment and comparator, and outcomes, for example. While it is likely that effectiveness trials will have more potential adopters of the intervention tested than do explanatory trials, both have a potential audience, and its members can only benefit from

the efforts by the investigators if these researchers pay attention in their reports to external validity issues as much as they emphasize internal validity.

Second, with respect to *most* of the entries on the checklist, researchers do not need to *do* more—they only need to *report* more. Nearly all of the information to be reported already exists in the researchers' files (e.g. treatment protocols and a count of the number of clients/patients consented) or in their heads (for instance, alternative treatments for the same condition published since study start; subjective judgments as to the client/patient categories the treatment may be most suitable for). It is just a question of making the information available to potential readers, either in the always concise report submitted to a journal, or as supplemental information available on the journal's website or by request from the corresponding author.

However, there are a number of entries on the list that require researchers to *collect* information that they may not routinely gather. Some elements are relatively easy to amass, but human subjects protections may be the barrier—for example, where it concerns collecting information on *potential* subjects who never gave informed consent. Fairly sophisticated data collection schemes may need to be put in place for other components, such as information on the therapy preferences and expectations of patients/clients, or on the adjunctive treatments used by patients, with or without the treating clinicians' approval or even knowledge. All of these new data collection efforts have a cost, in terms of the researcher's time and the goodwill of patients and treating clinicians.

Those investigators who have a sincere interest in seeing their findings used may want to make the investment. Treweek and Zwarenstein (2009, p. 7)

expressed it this way: "Trialists should routinely ask themselves at the design stage of their trial 'Who are the people I expect to use the results of my trial and what can I do to make sure that these people will not be forced to dismiss my trial as irrelevant to them, their patients, or their healthcare systems?'".

The checklist is intended as an aid to researchers in designing their studies as well as in writing their reports, to provide reasonable assurance that they do not omit collecting or conveying

any information that a reader might expect. Applicability depends on the nature of the intervention—there should be a more extensive report for a complex intervention directed at participation restrictions than for a medication directed at a change at the impairment level. The various items do not need to be addressed in research papers in any particular order,

but all applicable items need to find a place in the report (most likely in the Methods section and at the beginning of the Results section of a traditional research article, with the possible exception of some of the more speculative issues better relegated to the Discussion section). Their combination and order should be dictated by disciplinary traditions and what makes most sense to an author. The level of detail also depends on a number of factors, not the least journal space. Authors should write with that limit in mind (Dijkers, 2009), and plan a careful division of material between published text, website-published supplemental materials (e.g. tables with detail of patient/client characteristics) and yet other information made available on request (for instance, a detailed treatment manual).

### **Concluding thoughts**

In the past the "generalizability" of research findings has been defined as the feasibility of extrapolating to other settings, populations, and outcomes. Most if not all authors discussing explanatory vs.

---

***Most if not all authors discussing explanatory vs. pragmatic trials, or more specifically issues of applicability, transferability and extrapolation, have implicitly assumed that generalizability is inherent in a study and its findings.***

pragmatic trials, or more specifically issues of applicability, transferability and extrapolation, have implicitly assumed that generalizability is inherent in a study and its findings. In the present paper the argument was made that, because we do not deal with a random sample of participants, clinicians, settings, intervention versions, etc., we cannot generalize from study findings to other situations on a *statistical* basis. The transfer from a study situation to clinical practice or health system implementation always has to be a leap of faith by the decision makers involved; which leap will be based on the *perceived* similarity between their situation and the one the researcher dealt with. It is the responsibility of the researcher to make available all information on the study, its methods, subjects, interventions and outcomes. The practitioners have to make the decision to adopt an intervention, and in what format, based on their knowledge of the local situation as well as their expertise honed in didactic learning and practical training.

The claim by Balas and Boren (2000) that the pipeline from first investigation of a clinical innovation to routine use in clinical practice is 17 years long and delivers only 14% of the innovations to health care practice has been quoted frequently, including by the Institute of Medicine (Committee on Quality of Health Care in America, 2001). The barriers to rapid dissemination and implementation of new ideas into healthcare are multiple. Green, Ottoson, Garcia, and Hiatt (2009) mention that the blame for lack of rapid clinical application has been laid at the feet of “tradition-bound practitioners, who insist on practicing their way and believe they know their patients or populations best,” and at the smugness of “scientists believing that

if they publish it, practitioners and the public will use it” (p. 154). The reasons for the slow and limited uptake by practitioners of what does get published have been explored in a number of studies. The fact that researchers tend to write for other researchers, or at least fail to include in their reports those elements that decision makers need, is one of them. The checklist presented in this article is expected to help with that problem. If researchers are convinced that their findings are of immediate applicability in clinical practice, they should submit the research report

---

***If researchers are convinced that their findings are of immediate applicability in clinical practice, they should submit the research report to a journal that is read by clinicians, and should use the checklist to maximize the chance that the report will be sufficient to support translation of the research into practice.***

to a journal that is read by clinicians, and should use the checklist to maximize the chance that the report will be sufficient to support translation of the research into practice.

Many rehabilitation researchers, faced with the requirement to produce research of rigor and relevance, may have equated

rigor with internal validity, and possibly also with the psychometric quality of the outcome measures used, while relevance was understood to refer to the nature of the matter studied, and its importance to clients/patients and clinicians *per se*. However, they should consider that rigor and relevance do not coincide with internal vs. external validity, and that the two dimensions are fairly independent from one another. A study without any relevance to the needs of rehabilitation consumers or clinicians can be conducted with high internal validity—but so can one with high relevance.

The translation of research findings to practical applications requires as much prior rigor on the part of the reporting investigator, and one of those rigors is the complete and accurate collecting and reporting of information on those elements of one’s research design, implementation and findings that are needed



by clinicians and other practitioners, and clients/patients, to make decisions on the adoption of the intervention or innovation studied. The checklist provided here may serve as a guideline for achieving more rigor in reporting. There is no claim that accurate and complete reporting is *sufficient* to achieve knowledge translation, but it is a *necessary* first step for all research efforts initiated by a researcher—rather than ordered by a clinical organization.

In evaluating proposals for intervention research as well as manuscripts submitted to journals, peer reviewers are apt to focus on issues of internal validity, and tend to disregard external validity. Given that those aspects of design that concern internal validity are easier to evaluate (and reject as insufficient), this is not surprising. However, they should realize that accepting or rejecting research grant proposals and manuscripts for publication based primarily or exclusively on the adequacy of the study's internal validity will only serve to drive out proposals and reports of a pragmatic nature; only tightly controlled explanatory studies will survive the gauntlet of challenges. Although the surviving proposals and reports may be satisfactory to researchers, they will have little to offer to practitioners. The editors of journals that have as their mission improving rehabilitation, rather than building basic science, may want to put in place review criteria that allow for a balanced evaluation

of both the internal and the external validity of the research described (Green, Glasgow, Atkins, & Stange, 2009).

### **Postscript**

In the period between writing of this article and publication of the *FOCUS* issue, a number of papers have been published which address issues of external validity, using such terms as “directness,” “applicability” and “generalizability.” Although the issue of the applicability of evidence was originally brought up when the GRADE approach to rating the quality of evidence was first published (Guyatt et al., 2008), only recently have guidelines been published for downgrading evidence because it is indirect. GRADE distinguishes three types of indirectness: different population (e.g. children rather than adults); different intervention (e.g. telephone counseling instead of face-to-face contact); and different outcome measures (e.g. surrogate outcomes instead of patient-important outcomes) (Guyatt et al., 2011). The FORM approach asks for a rating of generalizability based on similarity of population, setting, disease stage and duration of illness (Hillier et al., 2011). While these papers concentrate on systematic reviewing and guideline development rather than the issues an individual practitioner deals with, and are medicine-focused, they are worthwhile reading for rehabilitation researchers.

## References

- Ahmad, N., Boutron, I., Moher, D., Pitrou, I., Roy, C., & Ravaud, P. (2009). Neglected external validity in reports of randomized trials: the example of hip and knee osteoarthritis. *Arthritis and Rheumatism*, 61(3), 361–369.
- Bagshaw, S. M., & Bellomo, R. (2008). The need to reform our assessment of evidence from clinical trials: a commentary. *Philosophy, Ethics, and Humanities in Medicine*, 3, 23.
- Balas, E. A., & Boren, S. A. (2000). Managing clinical knowledge for health care improvement. In J. H. van Bemmel & A. T. McCray (Eds.), *Yearbook of Medical Informatics 2000: Patient-centered Systems* (pp. 65–70). Stuttgart, Germany: Schattauer.
- Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K. F., Simel, D., & Stroup, D. F. (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA: The Journal of the American Medical Association*, 276(8), 637–639.
- Bonell, C., Oakley, A., Hargreaves, J., Strange, V., & Rees, R. (2006). Assessment of generalisability in trials of health interventions: suggested framework and systematic review. *BMJ (Clinical Research Ed.)*, 333(7563), 346–349.
- Bornhoft, G., Maxison-Bergemann, S., Wolf, U., Kienle, G. S., Michalsen, A., Vollmar, H. C., Gilbertson, S., & Matthiessen, P. F. (2006). Checklist for the qualitative evaluation of clinical studies with particular focus on external validity and model validity. *BMC Medical Research Methodology*, 6, 56.
- Boutron, I., Moher, D., Altman, D. G., Schulz, K. F., & Ravaud, P. for the CONSORT Group. (2008). Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Annals of Internal Medicine*, 148(4), 295–309.
- Britton, A., McKee, M., Black, N., McPherson, K., Sanderson, C., & Bain, C. (1999). Threats to applicability of randomised trials: exclusions and selective participation. *Journal of Health Services Research & Policy*, 4(2), 112–121.
- Burke, D. T., DeVito, M. C., Schneider, J. C., Julien, S., & Judelson, A. L. (2004). Reading habits of physical medicine and rehabilitation resident physicians. *American Journal of Physical Medicine & Rehabilitation*, 83(7), 551–559.
- Burke, D. T., Judelson, A. L., Schneider, J. C., DeVito, M. C., & Latta, D. (2002). Reading habits of practicing physiatrists. *American Journal of Physical Medicine & Rehabilitation*, 81(10), 779–787.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand-McNally.
- Centre for Evidence-Based Physiotherapy. (2011). PEDro: Physiotherapy Evidence Database. Retrieved from <http://www.pedro.org.au/>
- The Cochrane Collaboration - Cochrane Reviews*. (2010). Retrieved May 11, 2010, from <http://www.cochrane.org/cochrane-reviews>
- Committee on Quality of Health Care in America, Institute Of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, DC: National Academy Press.
- Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ (Clinical Research Ed.)*, 337, a1655. doi: 10.1136/bmj.a1655
- Davidoff, F., Batalden, P., Stevens, D., Ogrinc, G., & Mooney, S. E., for the SQUIRE development group. (2009). Publication guidelines for quality improvement studies in health care: evolution of the SQUIRE project. *BMJ (Clinical Research Ed.)*, 338, a3152. doi: 10.1136/bmj.a3152
- Davidson, K. W., Goldstein, M., Kaplan, R. M., Kaufmann, P. G., Knatterud, G. L., Orleans, C. T., Spring, B., Trudeau, K. J., & Whitlock, E. P. (2003). Evidence-based behavioral medicine: What is it and how do we achieve it? *Annals of Behavioral Medicine*, 26(3), 161–171.

- Dekkers, O. M., von Elm, E., Algra, A., Romijn, J. A., & Vandembroucke, J. P. (2010). How to assess the external validity of therapeutic trials: a conceptual approach. *International Journal of Epidemiology*, *39*, 89–94. doi: 10.1093/ije/dyp174
- Depp, C., & Lebowitz, B. D. (2007). Clinical trials: bridging the gap between efficacy and effectiveness. *International Review of Psychiatry*, *19*(5), 531–539.
- Des Jarlais, D. C., Lyles, C., Crepaz, N., & the TREND Group. (2004). Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *American Journal of Public Health*, *94*(3), 361–366.
- Dijkers, M. P. (2009). Ensuring inclusion of research reports in systematic reviews. *Archives of Physical Medicine and Rehabilitation*, *90*(11 suppl), S60–S69.
- Dijkers, M., Kropp, G. C., Esper, R. M., Yavuzer, G., Cullen, N., & Bakdalieh, Y. (2002). Quality of intervention research reporting in medical rehabilitation journals. *American Journal of Physical Medicine & Rehabilitation*, *81*(1), 21–33.
- Dzewaltowski, D. A., Estabrooks, P. A., Klesges, L. M., Bull, S., & Glasgow, R. E. (2004). Behavior change intervention research in community settings: how generalizable are the results? *Health Promotion International*, *19*(2), 235–245.
- Edlund, W., Gronseth, G., So, Y., & Franklin, G. (2004). *Clinical practice guidelines process manual –2004 edition*. St. Paul, MN: American Academy of Neurology.
- Faulkner, C., Fidler, F., & Cumming, G. (2008). The value of RCT evidence depends on the quality of statistical analysis. *Behaviour Research and Therapy*, *46*(2), 270–281.
- Gartlehner, G., Hansen, R. A., Nissman, D., Lohr, K. N., & Carey, T. S. (2006). A simple and valid tool distinguished efficacy from effectiveness studies. *Journal of Clinical Epidemiology*, *59*(10), 1040–1048.
- Gartlehner, G., Thieda, P., Hansen, R. A., Morgan, L. C., Shumate, J. A., & Nissman, D. B. (2009). Inadequate reporting of trials compromises the applicability of systematic reviews. *International Journal of Technology Assessment in Health Care*, *25*(3), 323–330.
- Glasgow, R. E., Bull, S. S., Gillette, C., Klesges, L. M., & Dziewaltowski, D. A. (2002). Behavior change intervention research in healthcare settings: a review of recent reports with emphasis on external validity. *American Journal of Preventive Medicine*, *23*(1), 62–69.
- Glasgow, R. E., Davidson, K. W., Dobkin, P. L., Ockene, J., & Spring, B. (2006). Practical behavioral trials to advance evidence-based behavioral medicine. *Annals of Behavioral Medicine*, *31*(1), 5–13.
- Glasgow, R. E., Klesges, L. M., Dziewaltowski, D. A., Estabrooks, P. A., & Vogt, T. M. (2006). Evaluating the impact of health promotion programs: using the RE-AIM framework to form summary measures for decision making involving complex issues. *Health Education Research*, *21*(5), 688–694.
- Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *American Journal of Public Health*, *89*(9), 1322–1327.
- Glasziou, P., Meats, E., Heneghan, C., & Shepperd, S. (2008). What is missing from descriptions of treatment in trials and reviews? *BMJ (Clinical Research Ed.)*, *336*(7659), 1472–1474. doi: 10.1136/bmj.39590.732037.47
- Green, L. W., Glasgow, R. E., Atkins, D., & Stange, K. (2009). Making evidence from research more relevant, useful, and actionable in policy, program planning, and practice slips “twixt cup and lip”. *American Journal of Preventive Medicine*, *37*(6 Suppl 1), S187–91. doi:10.1016/j.amepre.2009.08.017
- Green, L. W., Ottoson, J. M., Garcia, C., & Hiatt, R. A. (2009). Diffusion theory and knowledge dissemination, utilization, and integration in public health. *Annual Review of Public Health*, *30*, 151–174. doi: 10.1146/annurev.publhealth.031308.100049

- Guyatt, G. H., Oxman, A. D., Kunz, R., Vist, G. E., Falck-Ytter, Y., & Schunemann, H. J. for the GRADE Working Group. (2008). GRADE: What is "quality of evidence" and why is it important to clinicians? *BMJ (Clinical Research Ed.)*, 336(7651), 995–998. doi:10.1136/bmj.39490.551019.BE
- Guyatt, G. H., Oxman, A. D., Kunz, R., Woodcock, J., Brozek, J., Helfand, M., Alonso-Coello, P., Falck-Ytter, Y., Jaeschke, R., Vist, G., Akl, E. A., Post, P. N., Norris, S., Meerpohl, J., Shukla, V. K., Nasser, M., Schunemann, H. J., & The GRADE Working Group. (2011). GRADE guidelines: 8. Rating the quality of evidence—indirectness. *Journal of Clinical Epidemiology*, doi:10.1016/j.jclinepi.2011.04.014
- Hart, T., Fann, J. R., & Novack, T. A. (2008). The dilemma of the control condition in experience-based cognitive and behavioural treatment research. *Neuropsychological Rehabilitation*, 18(1), 1–21.
- Hillier, S., Grimmer-Somers, K., Merlin, T., Middleton, P., Salisbury, J., Tooher, R., & Weston, A. (2011). FORM: An Australian method for formulating and grading recommendations in evidence-based clinical guidelines. *BMC Medical Research Methodology*, 11, 23. doi:10.1186/1471-2288-11-23
- Jacquier, I., Boutron, I., Moher, D., Roy, C., & Ravaud, P. (2006). The reporting of randomized clinical trials using a surgical intervention is in need of immediate improvement: a systematic review. *Annals of Surgery*, 244(5), 677–683.
- Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J., Gavaghan, D. J., & McQuay, H. J. (1996). Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled Clinical Trials*, 17(1), 1–12.
- Jette, D. U., Bacon, K., Batty, C., Carlson, M., Ferland, A., Hemingway, R. D., Hill, J. C., Ogilvie, L., & Volk, D. (2003). Evidence-based practice: beliefs, attitudes, knowledge, and behaviors of physical therapists. *Physical Therapy*, 83(9), 786–805.
- Johnston, M. V., Sherer, M., & Whyte, J. (2006). Applying evidence standards to rehabilitation research. *American Journal of Physical Medicine & Rehabilitations*, 85(4), 292–309.
- Kendrick, T., Hegarty, K., & Glasziou, P. (2008). Interpreting research findings to guide treatment in practice. *BMJ (Clinical Research Ed.)*, 337, a1499. doi: 10.1136/bmj.a1499
- Klesges, L. M., Dzewaltowski, D. A., & Glasgow, R. E. (2008). Review of external validity reporting in childhood obesity prevention research. *American Journal of Preventive Medicine*, 34(3), 216–223.
- Klesges, L. M., Estabrooks, P. A., Dzewaltowski, D. A., Bull, S. S., & Glasgow, R. E. (2005). Beginning with the application in mind: designing and planning health behavior change interventions to enhance dissemination. *Annals of Behavioral Medicine*, 29 Suppl, 66–75.
- Leeman, J., Jackson, B., & Sandelowski, M. (2006). An evaluation of how well research reports facilitate the use of findings in practice. *Journal of Nursing Scholarship*, 38(2), 171–177.
- Maher, C. G., Sherrington, C., Herbert, R. D., Moseley, A. M., & Elkins, M. (2003). Reliability of the PEDro scale for rating quality of randomized controlled trials. *Physical Therapy*, 83(8), 713–721.
- Marks, D. F. (2009). How should psychology interventions be reported? *Journal of Health Psychology*, 14(4), 475–489.
- Marks, D. F. (2010). Publication guidelines for intervention studies in the Journal of Health Psychology. *Journal of Health Psychology*, 15(1), 5–7.
- Moher, D., Schulz, K. F., & Altman, D. for the CONSORT Group (Consolidated Standards of Reporting Trials). (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA: The Journal of the American Medical Association*, 285(15), 1987–1991. doi: 10.1001/jama.285.15.1987
- Mohr, D. C., Spring, B., Freedland, K. E., Beckner, V., Arean, P., Hollon, S. D., Ockene, J., & Kaplan, R. (2009). The selection and design of control conditions for randomized controlled trials of

psychological interventions. *Psychotherapy and Psychosomatics*, 78(5), 275–284.

Ogrinc, G., Mooney, S. E., Estrada, C., Foster, T., Goldmann, D., Hall, L. W., Huizinga, M. M., Liu, S. K., Mills, P., Neily, J., Nelson, W., Pronovost, P. J., Provost, L., Rubenstein, L. V., Speroff, T., Splaine, M., Thomson, R., Tomolo, A. M., & Watts, B. (2008). The SQUIRE (Standards for QUality Improvement Reporting Excellence) guidelines for quality improvement reporting: explanation and elaboration. *Quality & Safety in Health Care*, 17 Suppl 1, i13–32.

Oxford Centre for Evidence-based Medicine. (2009). *Oxford Centre for Evidence-based Medicine—Levels of Evidence (March 2009)*. Retrieved from <http://www.cebm.net/?o=1025>

Philibert, D. B., Snyder, P., Judd, D., & Windsor, M. M. (2003). Practitioners' reading patterns, attitudes, and use of research reported in occupational therapy journals. *The American Journal of Occupational Therapy*, 57(4), 450–458. doi: 10.5014/ajot.57.4.450

Pibouleau, L., Boutron, I., Reeves, B. C., Nizard, R., & Ravaud, P. (2009). Applicability and generalisability of published results of randomised controlled trials and non-randomised studies evaluating four orthopaedic procedures: methodological systematic review. *BMJ (Clinical Research Ed.)*, 339, b4538. doi: 10.1136/bmj.b4538

Powell, C. A., & Case-Smith, J. (2003). Information literacy skills of occupational therapy graduates: a survey of learning outcomes. *Journal of the Medical Library Association*, 91(4), 468–477.

Robb, S. L., & Carpenter, J. S. (2009). A review of music-based intervention reporting in pediatrics. *Journal of Health Psychology*, 14(4), 490–501. doi: 10.1177/1359105309103568

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. doi:10.1016/S0140-6736(04)17670-8

Shapiro, S. H., Weijer, C., & Freedman, B. (2000). Reporting the study populations of clinical trials. Clear transmission or static on the line? *Journal of Clinical Epidemiology*, 53(10), 973–979.

Tansella, M., Thornicroft, G., Barbui, C., Cipriani, A., & Saraceno, B. (2006). Seven criteria for improving effectiveness trials in psychiatry. *Psychological Medicine*, 36(5), 711–720.

Thorpe, K. E., Zwarenstein, M., Oxman, A. D., Treweek, S., Furberg, C. D., Altman, D. G., Tunis, S., Bergel, E., Harvey, I., Magid, D. J., & Chalkidou, K. (2009). A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *Journal of Clinical Epidemiology*, 62(5), 464–475.

Treweek, S., & Zwarenstein, M. (2009). Making trials matter: pragmatic and explanatory trials and the problem of applicability. *Trials*, 10, 37. doi:10.1186/1745-6215-10-37

Turner, P., & Whitfield, A. (1997). Journal readership amongst Australian physiotherapists: a cross-national replication. *Australian Journal of Physiotherapy*, 43(3), 197–200.

Vallvé, C. (2003). A critical review of the pragmatic clinical trial. [Revisión crítica del ensayo clínico pragmático]. *Medicina Clínica*, 121(10), 384–388.

Van Spall, H. G., Toren, A., Kiss, A., & Fowler, R. A. (2007). Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *JAMA: The Journal of the American Medical Association*, 297(11), 1233–1240. doi: 10.1001/jama.297.11.1233

Zwarenstein, M., Treweek, S., Gagnier, J. J., Altman, D. G., Tunis, S., Haynes, B., Oxman, A. D., & Moher, D. for the CONSORT and Pragmatic Trials in Healthcare (Practihc) groups. (2008). Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ (Clinical Research Ed.)*, 337, a2390. doi: 10.1136/bmj.a2390



## Recommended Citation

Dijkers, M. P. J. M. (2011). External validity in research on rehabilitative interventions: Issues for knowledge translation. *FOCUS Technical Brief* (33). Austin, TX: SEDL, National Center for the Dissemination of Disability Research

## Address for Correspondence

Marcel Dijkers, PhD  
Department of Rehabilitation Medicine, Box 1240  
Mount Sinai School of Medicine  
One Gustave L. Levy Place  
New York, NY 10029-6574  
Telephone: [00-1-] 212-659-8587  
Fax: [00-1-] 212-348-5901  
e-mail: marcel.dijkers@mssm.edu

## Author

**Marcel Dijkers, PhD, FACRM**, is Research Professor at Mount Sinai School of Medicine, Department of Rehabilitation Medicine and is a senior investigator on the NIDRR-funded New York TBI and SCI Model System centers at Mount Sinai. He was facilitator of the NCDDR's Task Force on Systematic Review and Guidelines from 2006 to 2011. Dr. Dijkers is a former President of the American Congress of Rehabilitation Medicine. He sits on the editorial board of the *Journal of Head Trauma Rehabilitation* and the *Journal of Physical Medicine and Rehabilitation Sciences*, and is an editor of *Rehabilitation Research and Practice*.

Copyright © 2011 by SEDL

*FOCUS: A Technical Brief From the National Center for the Dissemination of Disability Research* was produced by the National Center for the Dissemination of Disability Research (NCDDR) under grant H133A060028 from the National Institute on Disability and Rehabilitation Research (NIDRR) in the U.S. Department of Education's Office of Special Education and Rehabilitative Services (OSERS). NCDDR is funded 100% by NIDRR at \$750,000 per project year.

## NATIONAL CENTER FOR THE DISSEMINATION OF DISABILITY RESEARCH

Advancing Research, Improving Education



NCDDR's scope of work focuses on developing systems for applying rigorous standards of evidence in describing, assessing, and disseminating outcomes from research and development sponsored by NIDRR. The NCDDR promotes movement of disability research results into evidence-based instruments such as systematic reviews as well as consumer-oriented information systems and applications.

SEDL operates the NCDDR. SEDL is an equal employment opportunity/affirmative action employer and is committed to affording equal employment opportunities for all individuals in all employment matters. Neither SEDL nor the NCDDR discriminates on the basis of age, sex, race, color, creed, religion, national origin, sexual orientation, marital or veteran status, or the presence of a disability. The contents of this document do not necessarily represent the policy of the U.S. Department of Education, and you should not assume endorsement by the federal government.