[17] Meier C, Woitge HW, Witte K, Lemmer B, Seibel MJ. Supplementation with oral vitamin D3 and calcium during winter prevents seasonal bone loss: a randomized controlled open-label prospective trial. *J Bone Miner Res* 2004;**19:**1221–30.

[18] Travers J, Marsh S, Williams M *et al*. External validity of randomised controlled trials in asthma: to whom do the results of the trials apply? *Thorax* 2007;**62:**219–23.

[19] Travers J, Marsh S, Caldwell B *et al*. External validity of randomized controlled trials in COPD. *Respir Med* 2007;**101:**1313–20.

[20] Doll R, Hill AB. Lung cancer and other causes of death in relation to smoking; a second report on the mortality of British doctors. *Br Med J* 1956;**2:**1071–81.

[21] Weiss NS, Koepsell TD, Psaty BM. Generalizability of the results of randomized trials. *Arch Intern Med* 2008;**168:**133–35.

[22] Pitt B, Zannad F, Remme WJ *et al*. The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomized Aldactone Evaluation Study Investigators. *N Engl J Med* 1999;**341:** 709–17.

[23] Juurlink DN, Mamdani MM, Lee DS *et al*. Rates of hyperkalemia after publication of the Randomized Aldactone Evaluation Study. *N Engl J Med* 2004;**351:** 543–51.

[24] Rothman KJ, Greenland S. *Modern Epidemiology*. Philadelphia: Lippincott Williams & Wilkins, 1998.

Published by Oxford University Press on behalf of the International Epidemiological Association

International Journal of Epidemiology 2010;**39**:94–96
doi:10.1093/ije/dyp305

Commentary: External validity of results of randomized trials: disentangling a complex concept

Peter M Rothwell

University Department of Clinical Neurology, Level 6, West Wing, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK. E-mail: peter.rothwell@clneuro.ox.ac.uk

Accepted 25 August 2009

It is now widely accepted that in most situations, randomized controlled trials (RCTs) and systematic reviews are the most reliable methods of determining the effects of treatment. Yet, the methodology is still relatively new (a few decades old), and so our understanding of trial design, and more especially of how best to make use of results, is less than perfect. RCTs must be internally valid (i.e. their design and conduct must minimize the possibility of bias), and, until recently, guidelines on trial methodology and reporting, such as the CONSORT initiative, concentrated almost completely on issues related to internal validity. However, to be clinically useful, the result of a trial must also be relevant to clinical practice, i.e. be reasonably likely to be replicated when applied to a definable group of patients in a particular clinical setting. The extent to which a result can be extrapolated in this way has been variously termed as 'external validity, applicability or generalizability'.[1]

For some interventions, such as lowering blood pressure in chronic uncontrolled hypertension, the benefits have been shown to be generalizable to the vast majority of patients and settings, but the effects of other interventions will often depend on factors such as the characteristics of the patient, the method of application of the intervention and the setting of treatment. How these factors are taken into account in the design and performance of an RCT and in the reporting of the results can have a major impact on the clinical usefulness of the result. Lack of external validity has always been the most frequent criticism by clinicians of RCTs, systematic reviews and guidelines. Although much more research is required, systematic assessments of the external validity of trials in specific areas of medicine are now beginning to demonstrate the often substantial disparity between the information that is provided by RCTs and the information that is actually required by clinicians.[2,3] This disparity is one explanation for the underuse in routine practice of many treatments that have been shown to be beneficial in trials and are recommended in guidelines.

However, external validity is a 'slippery' concept. It can be defined in broad terms, as above, but is much more difficult to quantify exactly. While the determinants of internal validity are intuitive and

Table 1 Issues that potentially affect external validity and which should be addressed in reports of the results of RCTs and considered by clinicians[1]

- **Setting of the trial**
  - Health care system
  - Country
  - Recruitment from primary, secondary or tertiary care
  - Selection of participating centres
  - Selection of participating clinicians
- **Selection of patients**
  - Methods of pre-randomization diagnosis and investigation
  - Eligibility criteria
  - Exclusion criteria
  - Placebo run-in period
  - Treatment run-in period
  - 'Enrichment' strategies
  - Ratio of randomized patients to eligible non-randomized patients in participating centres
  - Proportion of patients who declined randomization
- **Characteristics of randomized patients**
  - Baseline clinical characteristics
  - Racial group
  - Uniformity of underlying pathology
  - Stage in the natural history of their disease
  - Severity of disease
  - Comorbidity
  - Absolute risks of a poor outcome in the control group
- **Differences between the trial protocol and routine practice**
  - Trial intervention
  - Timing of treatment
  - Appropriateness/relevance of control intervention
  - Adequacy of non-trial treatment—both intended and actual
  - Prohibition of certain non-trial treatments
  - Therapeutic or diagnostic advances since trial was performed
- **Outcome measures and follow-up**
  - Clinical relevance of surrogate outcomes
  - Clinical relevance, validity and reproducibility of complex scales
  - Effect of intervention on most relevant components of composite outcomes
  - Who measured outcome
  - Use of patient-centred outcomes
  - Frequency of follow-up
  - Adequacy of the length of follow-up
- **Adverse effects of treatment**
  - Completeness of reporting of relevant adverse effects
  - Rates of discontinuation of treatment
  - Selection of trial centres and/or clinicians on the basis of skill or experience
  - Exclusion of patients at risk of complications
  - Exclusion of patients who experienced adverse effects during a run-in period
  - Intensity of trial safety procedures

can therefore be worked out from first principles and quality scores developed, understanding of the determinants of the external validity of an RCT requires clinical rather than statistical expertise and usually depends on a detailed understanding of the particular clinical condition under study and its management in routine clinical practice. External validity is also highly dependent on the particular perspective of the individual making the judgement. For one clinician with a particular patient, a trial result might be almost perfectly applicable, whereas for another clinician and patient the external validity may be extremely low.

Some of the main potential determinants of external validity have been highlighted in a previous review (Table 1).[1] In the accompanying discussion paper, Dekkers and colleagues[4] attempt to subclassify some of these determinants of external validity. They distinguish two concepts: (i) whether the results of a trial are valid for patients other than those in the original study population in a treatment setting that is in all respects equal to the treatment setting in the original study ('external validity'); (ii) whether the results are valid for patients to whom they are generalizable but who are in a different treatment setting than the original study population ('applicability').

These two concepts are distinct and can be used to subclassify some of the determinants of external validity in Table 1. The subsections of the table on the setting of the trial and on differences between trial protocol and routine practice cover all of the issues identified by Dekkers and colleagues as relating to 'applicability', and the subsections on selection of patients and characteristics of randomized patients cover 'external validity'. However, there is a third concept that is apparent in the table, which is not covered by these two concepts; i.e. those aspects of trial design and performance that influence the generalizability of the result irrespective of the patient population or the clinical setting. The subsection of the table on outcome measures and follow-up and some of the issues under adverse effects and outcomes are often crucial to the clinical usefulness of a trial result.

On balance, it is perhaps better to agree on the term 'external validity' as the overarching descriptor for all aspects of the design and performance that impact on the external usefulness of the result of a trial, independent of the 'internal validity' of the trial. Whether external validity should then be subclassified into pragmatic groupings, as in the table, into conceptual groupings, as advocated by Dekker and colleagues, or perhaps both, needs to be debated. For example, another conceptual distinction could be (i) those aspects of design, performance and reporting of a trial that are likely to affect external validity from the perspectives of all patients and clinicians in all settings (e.g. the subsection in the table on outcome measures and follow-up); or (ii) those aspects of design, performance and reporting of a trial that are likely to affect external validity from the perspectives of some patients, clinicians and settings but not others (e.g. the subsections on setting, selection of patients and characteristics of patients).

This classification would have the advantage of separating out the aspect of external validity that was more consistently quantifiable from that which depended on the particular perspective of the observer.

Irrespective of what classification is eventually decided to be best (or least imperfect), Dekker and colleagues' contribution is certainly helpful in highlighting the need for greater consideration of external validity of randomized trials in general, such that that treatments are used appropriately in as many patients as possible in routine clinical practice. Finally, the results of a single trial will rarely be relevant to all patients and all settings, but it is important that clinicians do not exaggerate the extent of the problem. The results of trials should be assumed to be externally valid unless there are specific reasons to put this assumption into significant doubt. Even then, the trial result may still be the most reliable, available estimate of the likely effect of treatment in a particular patient in a particular setting. Perhaps the greatest current problem is, therefore, that trials are often not reported in sufficient detail to allow clinicians to judge to whom the results can reasonably be applied.

Conflict of interest: None declared.

References

[1] Rothwell PM. External validity of randomised controlled trials: to whom do the results of this trial apply? *Lancet* 2005;**365:**82–93.

[2] Travers J, Marsh S, Williams M *et al*. External validity of randomised controlled trials in asthma: to whom do the results of the trials apply? *Thorax* 2007;**62:**219–23.

[3] Petersen MK, Andersen KV, Andersen NT, Sballe K. ''To whom do the results of this trial apply?'' External validity of a randomized controlled trial involving 130 patients scheduled for primary total hip replacement. *Acta Orthop* 2007;**78:**12–18.

[4] Dekkers OM, von Elm E, Algra A, Romijn JA, Vandenbroucke JP. How to assess the external validity of therapeutic trials: a conceptual approach. *Int J Epidemiol* 2010;**39:**89–94.