

# Reading tips for the clinician 3

## Are study results good enough to be generalised?

With more and more information becoming available to physical therapists, much of it on the Internet, it is increasingly important to be selective in what you read. In this, the third in a series of three Keynotes, Joan M Walker continues her step-by-step approach to assessing how useful an article will be to your practice.

The first paper looked at the value of review articles, how to assess their quality, and how to select useful papers simply by reading the abstract. The second looked in more detail at the criteria a physical therapist should use when determining whether a review article or original research is worth reading. This final paper provides advice on how to assess the statistical validity of a study.

You have a question relating to your clinical practice that needs answering. You locate several papers. The titles are relevant and interesting, the authors have a good track record, the abstracts reveal issues pertinent to your question, and the intent of the papers is on target. The next level of questions you need to pose to refine your reading should relate to the studies' samples.

### Questions to ask about samples

1. Was the sample randomly selected in an appropriate manner?
2. Was the sample size justified?
3. If the sample was not selected

randomly, was the selection process justified?

4. Was selection bias avoided?
5. Was the sample described in a way that would allow it to be duplicated? Are sufficient details given on inclusion and exclusion criteria?
6. Were subjects randomly allocated into groups?
7. If the subjects were not allocated into groups randomly, was the way the groups were allocated justified?
8. Were all subjects admitted into the study accounted for? At least 80% that started should be reported in outcomes analysis.
9. Were inclusion criteria clear and compatible for all studies included in a review article?
10. Is the sample like your clinical population, and is your clinical facility similar to the study site?

### Differences between study topics

Some questions are particularly important to ask, but these vary according to the types of articles you are searching for.

- In articles about therapy question 1 is most important.
- In articles about diagnostic tests Ask: "Did the patient sample include an appropriate spectrum of patients to whom the diagnostic test should be applied in clinical practice?"
- In articles about harm Ask: "Were the groups being compared similar in all ways that might affect the outcome?" You want to be sure there weren't other factors, apart from the one being examined, that might have affected the result.
- In articles about prognosis Ask: "Was there a representative and

well-defined sample of patients at a similar point in the course of the disease?"

### Points to remember about sampling

When an investigator reports taking a random sample this should mean that anyone meeting inclusion criteria had an equal chance of selection.

There are several sampling techniques. In probability sampling every unit of the population should have a specified chance of selection. Non-probability sampling doesn't fit this criterion, and results cannot be inferred from this to the whole population. Many investigators report use of a sample of convenience, where the subjects are simply readily accessible - for example physical therapy students. But there usually is no evidence that such a sample is representative of age, gender and education population groups.

Whether a sample is representative or not affects the validity of a study.

### About validity

There are many types of validity.

#### ● Face validity

If you were looking at tools to assess low back pain, a tool would have face validity if it subjectively appeared to address the problem and others in the field agreed with that conclusion. It is the weakest form of validity but usually the first tested.

#### ● Construct validity

A tool to assess low back pain would have construct validity if there were clear operational definitions of variables associated with low back pain - for example, activities of daily living involving trunk motion.

## ● Criterion validity

A tool may have construct validity when the scores on a test are similar to that of the “gold standard”. For example, a new outcome measure for low back pain might be compared with the Oswestry Low Back Pain Disability Questionnaire. The important factor is that the reader accepts the latter tool as a “gold standard”.

## ● Concurrent validity

The extent to which two tests are comparable at the time of testing.

## ● Predictive validity

The ability of the tool to predict what it says it predicts. Did it do that and how well? Was sensitivity and specificity reported?

## ● Internal and external validity

As a critical reader, you want to establish if a study has internal and external validity.

*Internal validity* exists when the changes in the effect (dependent characteristic (variable) are due to the intervention (independent variable). Among the threats to internal validity are: therapist bias, imitation of treatment by patients, competitive rivalry or resentment in patients, group assignment, factors related to history, maturation, testing, patient instruction, mortality and statistical regression toward the mean.

*External validity* is the extent to which the results can be applied to a larger population than that tested. Among the threats to external validity are subject selection, the setting and time period in which the study was done. Randomised clinical trials have much stronger external validity than single case studies.

## Assessing validity in primary studies

If you want to know whether therapy results are valid, and you only review primary studies (first report of data), essential questions to pose are:

- Was the assignment into study groups randomised?
- Were all patients (subjects) who entered the trial properly accounted for at the conclusion?
- If more than one treatment was given to every patient, was the

order of treatments (interventions) randomised? If the number of interventions was two, the first intervention received should have been randomised. If the number exceeds two, then the investigator should have used a technique to randomise the intervention order.

- Was the follow-up complete?
- How many, if any, were lost to follow-up? Were reasons for this loss given?
- Were all patients analysed in the groups to which they were randomised?

Accounting for all patients at the end of a trial, and noting totals on follow-ups, is important as those that were lost may differ in some substantial manner from those that remain. The reader can decide this by assuming all who were lost from the treatment group did badly while those lost from the control group did well. If this makes no difference to the conclusion, then the loss to follow-up was not excessive.

Ensuring that patients were analysed in the group they were assigned is important because it preserves randomisation. Keeping patients in their original groups helps to ensure that factors that are known and factors that are not known (to the investigator) will be equally distributed, and that any effect observed should be due to the treatment assigned.

## Assessing validity in secondary guides

If you are reviewing secondary guides (review articles, textbooks) to therapy, the following questions should be posed for all studies included in the review:

- Were patients, clinicians, study personnel masked to the intervention?
- Where patients and clinicians could not be kept masked, were steps taken to minimise bias by masking the assessor(s)?
- Did the persons giving the treatments also assess the outcomes? If they did, there is a

danger of bias

- Were groups similar at the start of the study?
- Apart from the experimental intervention, were groups treated the same? If there was a second intervention, was it given to both groups?
- Were studies compared similar in research design, sample sizes, time period and equipment used?

## In summary

Abandon the article and go to the next on your list if the sample selection was unclear, the method of sampling inappropriate, the sample inappropriate for the question posed, or if the sample is not similar to your clinical population. “No” answers to the above questions about validity should cause you to go to the next article. Use of a set of structured questions consistently will assist you to be an efficient, critical reader.

## Further reading

Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group: Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? *JAMA* 1993;270(21):2598-2601.

Helewa A, Walker JM: *Critical Evaluation of Research in Physical Rehabilitation. Towards Evidence-Based Practice.* Philadelphia, W.B. Saunders, 2000, ISBN 0-7216-7390-2

Hulley SB, Cummings SR: *Designing Clinical Research. An Epidemiologic Approach.* Baltimore, Williams & Wilkins, 1988.

Joan M Walker, PhD, DipTP, BPT, FAPTA, FNZSP professor Emeritus at Dalhousie University, Canada, is a world-renowned researcher and writer. In 1999 she received WCPT's greatest honour, the Mildred Elson Award, for her contributions to the physical therapy profession.

*Keynotes is a series of occasional papers dealing with important professional, practice and policy issues relevant to physical therapists across the world.*

*They are written by independent authors and do not necessarily represent WCPT's opinion. For further information contact: WCPT*

*Kensington Charity Centre  
4th Floor Charles House  
375 Kensington High Street  
London W14 8QH, UK  
Tel: 020 7471 6765  
Email: [info@wcpt.org](mailto:info@wcpt.org)  
[www.wcpt.org](http://www.wcpt.org)*

WCPT is a registered charity in the UK, no 234307  
© WCPT 2005