

# Outcome measures measure outcomes, not effects of intervention

Robert Herbert<sup>1</sup>, Gro Jamtvedt<sup>2</sup>, Judy Mead<sup>3</sup> and Kåre Birger Hagen<sup>4</sup>

<sup>1</sup>University of Sydney, Australia <sup>2</sup>Norwegian Centre for Health Services Research, Norway <sup>3</sup>The Chartered Society of Physiotherapy, UK <sup>4</sup>National Resource Centre for Rehabilitation in Rheumatology, Diakonhjemmet Hospital, Norway

Historically, measurement of clinical outcomes was not a feature of routine clinical practice. Physiotherapists (and, for that matter, most other health professionals) did not systematically collect data on patients' outcomes. Typically physiotherapists obtained incidental impressions of clinical outcomes from clinical observations, or from patients' comments about satisfaction (or dissatisfaction) with physiotherapy services.

In more recent times there has been pressure on physiotherapists to become more accountable for their practices. The pressure has come from makers of health care policies, those who allocate and fund health care (government, insurers, managers), and from within the physiotherapy profession. One of the driving forces has been the perception that physiotherapists must justify what they do. It is thought that by providing evidence of good clinical outcomes physiotherapists can demonstrate that what they do is worthwhile.

In the last two decades the physiotherapy profession has taken up the call for more and better clinical measurement. An early landmark was the publication, in 1985, of *Measurement in Physical Therapy* (Rothstein 1985). More recently there has been a proliferation of textbooks, journal features, and web sites documenting clinical outcome measures and their measurement properties (see, for example, Wade 1992, Koke et al 1999, Maher et al 2000 and Finch et al 2002). In some countries at least, a large proportion of physiotherapists routinely document clinical outcomes using validated tools. In New South Wales, WorkCover pays an additional fee to practitioners who adequately document measures of clinical outcomes.

Perhaps it is unfortunate that the physiotherapy profession has responded to the perception that physiotherapists must justify what they do by routinely measuring clinical outcomes. The implication is that measures of outcome can provide justification for intervention. Arguably that is not the case. Outcome measures measure outcomes. They do not measure the effects of intervention. Outcomes of interventions and effects of interventions are very different things.

Clinical outcomes are influenced by many factors other than intervention, including the natural course of the condition, statistical regression, placebo effects, and so on. (Tuttle (2005) makes this point clearly in his article, in this issue, on the predictive value of clinical outcome measures.) The implication is that a good outcome does not necessarily indicate that intervention was effective; the good outcome

may have occurred even without intervention. And a poor outcome does not necessarily indicate that intervention was ineffective; the outcome may have been worse still without intervention. This is why proponents of evidence-based physiotherapy, including ourselves (Herbert et al 2005), argue it is necessary to look to randomised trials to determine, with any degree of certainty, the effects of intervention. It is illogical, on the one hand, to look to randomised controlled trials for evidence of effects of interventions while, on the other hand, seeking justification for the effectiveness of clinical practice with uncontrolled measurement of clinical outcomes.

Taken further, this line of reasoning suggests that, at least in some circumstances, measures of individual patients' clinical outcomes should have no role in influencing decisions about treatment for those patients. According to this extreme view, randomised trials provide better information about the effects of intervention than do measures of clinical outcomes. So decisions about intervention for a particular patient should be based entirely on the findings of randomised trials, without regard to the apparent effects of treatment suggested by measures of clinical outcome on that patient. For example, if a randomised trial suggests that, on average, an intervention produces effects that a patient considers would be worthwhile, the implication is that intervention should continue to be offered even if the patient's outcomes are poor. The reasoning goes that the best we can know of the effects of a treatment (from randomised trials) tells us that this intervention typically produces clinically worthwhile effects. The patient may be one of the unlucky patients who does not benefit from (or is harmed by) this intervention, or it may be that the patient's poor outcomes might have been worse still without the intervention. We cannot discriminate between these scenarios so we act on the basis of what we think is most likely to be true: on average the intervention is helpful. Consequently we continue to provide the intervention even though the outcome of intervention is poor.

This view is completely antithetical to the empirical approach to clinical practice exemplified by some authors, and notably by Maitland in his classic texts on manual therapy (Maitland et al 2001). In the fully empirical approach, intervention is always followed by assessment of clinical outcomes. If outcomes improve, the intervention may be continued until the problem is completely resolved. If outcomes do not improve or worsen, the intervention is modified or discontinued. This approach appears to be reasonable, but it involves making clinical decisions on the basis of information that is very difficult to interpret. The empirical approach, in

which clinical decisions are based on careful measurement of outcomes, is not evidence-based physiotherapy. If we base clinical decisions about intervention on high quality clinical research, measures of clinical outcome need have little role in clinical decision-making or in justifying clinical practice. Interventions can be recommended without consideration of their outcomes.

Is there any role for clinical outcome measures in clinical decision making? We think that, when there is evidence of effects of intervention from high quality clinical trials, a sensible approach to clinical decision-making lies somewhere between the two extremes of the fully empirical approach and a hard-line approach in which clinical decision-making is based only on high quality clinical research without regard to outcome. In this approach, extreme clinical observations (very good or very poor outcomes) are considered likely to be 'real' (bias is unlikely to have qualitatively altered the clinical picture), so they are used to guide clinical decision-making. On the other hand, the qualitative interpretation of typical observations (small improvements in outcome) could plausibly be altered by bias, so they are essentially ignored. In other words, this approach suggests that clinical decision making should be influenced by observations of very good and very poor outcomes, but should not be influenced by less extreme observations.

What does this mean in practice? It means, first of all, that there is value in careful measurement of clinical outcomes, because extreme clinical outcomes influence clinical decision making. It also means that the degree of regard we pay to measures of clinical outcome depends on how extreme the outcomes are. When outcomes are *very* poor we *should* discontinue the intervention, even if the best clinical trials tell us that the intervention is, on average, effective, because a *very* poor outcome is unlikely to be explicable just by confounding effects such as the natural course of the condition, statistical regression, and so on — it probably also reflects that this person truly responded poorly to the intervention. On the other hand, less extreme poor outcomes might reasonably be ignored, and an intervention might be persisted with regardless of moderately poor outcome if the best clinical trials provide strong evidence that the intervention produces, on average, a clinically worthwhile effect. (Of course such clinical decisions should be informed by patients' preferences — for a variety of reasons some patients will want to continue with an intervention even when they have very poor outcomes, or discontinue an intervention even when the outcomes are excellent.)

Clinical outcome measures become more important when there is little or no evidence from high quality randomised trials. In that case, the alternatives to relying on clinical outcome measures are unattractive: we could choose not to intervene at all, or we could intervene in the absence of high quality evidence and use potentially misleading low quality

evidence to guide decisions about intervention. In contrast, when randomised trials provide clear evidence of the effects of an intervention from high quality clinical trials, clinical outcome measures become relatively unimportant and measures of the process of care become more useful. When evidence of effects of interventions is strong, we should audit the process of care to see if it is consistent with what the evidence suggests is good practice. When there is little or no evidence (i.e. when practice cannot be evidence-based) we should audit clinical outcomes.

In concluding, we note that there is another role for measurement of outcomes other than its limited role in telling us about the effects of intervention. Routine standardised outcome measurements can be used to generate practice-specific estimates of prognosis. For example, a physiotherapist who routinely assesses the presence or absence of shoulder pain in stroke patients at discharge following an upper limb rehabilitation program can use those data to generate practice-specific prognoses about the risk of developing shoulder pain by the time of discharge. In addition, as Hahne (2004) and Tuttle (2005) have shown, simple measures of clinical outcome can provide patient-specific prognostic information. It is important to recognise that these data have useful prognostic value, but they do not provide good evidence of the effectiveness or otherwise of intervention.

## References

- Finch E, Brooks D, Stratford P and Mayo N (2002): *Physical Rehabilitation Outcome Measures: A Guide to Enhanced Clinical Decision Making* (2nd edn). Philadelphia: Lippincott, Williams and Wilkins.
- Hahne AJ, Keating JL and Wilson SC (2004): Do within-session changes in pain intensity and range of motion predict between-session changes in patients with low back pain? *Australian Journal of Physiotherapy* 50:17–23.
- Herbert RD, Jamtvedt G, Mead J and Hagen KB (2005): *Practical Evidence-Based Physiotherapy*. Oxford: Elsevier.
- Koke AJA, Heuts PHTG, Vlaeyen JWS and Weber W (1999): *Meetinstrumenten Chronische Pijn. Deel 1 Functionele Status*. Maastricht: Pijn Kennis Centrum S.
- Maher C, Latimer J and Refshauge K (2000): *Atlas of Clinical Tests and Measures for Low Back Pain*. Melbourne: Australian Physiotherapy Association.
- Maitland GD, Hengeveld E, Banks K and English K (2001): *Maitland's Vertebral Manipulation*. Oxford: Butterworth-Heinemann.
- Rothstein J (ed) (1985): *Measurement in Physical Therapy*. New York: Churchill Livingstone.
- Tuttle N (2005): Do changes within a manual therapy treatment session predict between-session changes for patients with cervical spine pain? *Australian Journal of Physiotherapy* 51: 42–47.
- Wade DT (1992): *Measurement in Neurological Rehabilitation*. Oxford: Oxford University Press.