

## EDITORIAL

**Measurement and Evaluation of Outcomes in Rehabilitation**

For the management and evaluation of rehabilitation, outcome measures have become more and more essential. This is partly due to the increasing need for evidence-based medicine and service accreditation. There are a number of existing instruments, but many were developed before the awareness of modern psychometric methods that have been shown to be useful in clinical practice. The terminology developed within the International Classification of Functioning, Disability and Health (ICF) has become a valuable framework for developing and evaluating outcome measures through linkage to ICF category codes. In the symposium arranged by the *Journal of Rehabilitation Medicine* on September 27–28, 2004 in Stockholm, a number of related issues were addressed. The purpose of the symposium was to present an up-to-date review of methodological aspects of instruments and evaluation of outcomes in rehabilitation as exemplified by clinically used methods and instruments and discussion of criteria for appropriate choice of methods and instruments. We were fortunate to have a number of leading international scientists as invited speakers. There were also poster sessions. Almost 180 persons participated from 24 different countries; most of the participants were from Europe. In this editorial, we will highlight some of the main messages presented at the symposium.

*Allen Heinemann* from Chicago gave an overview of the concepts and structure of outcome measures with emphasis on quality indicator applications. He reminded us of the key components of healthcare quality measurement developed by Donabedian: Structure – can you get access to the system? Process – what happens to you in the system? Outcome – how well has the service been delivered? Process indicators may be satisfaction with available services and stakeholder satisfaction with information provided. A model for studying patient satisfaction with rehabilitation using a newly developed questionnaire was presented. Outcome indicators may be improved function, productivity and participation, and living situation. It was also pointed out that accrediting organizations are promoting performance indicators for rehabilitation.

*Alarcos Cieza* from Munich summarized the point that the ICF is a valuable framework to describe and evaluate rehabilitation. She collaborates with an international group to develop ICF core sets for different common medical conditions, as published recently in a supplement to *Journal of Rehabilitation Medicine* (1). Furthermore, different commonly used instruments are being linked to ICF, with ICF codes linked to the individual items in the instruments (2). Cieza clearly pointed out that ICF is not an instrument but a conceptual framework from which instruments can be developed.

The unfortunate non-critical use of ordinal data without awareness of their limited psychometric properties was ad-

ressed in several lectures. *Anne Fisher* from Umeå introduced the alternative use of Rasch measurement methods, an item response theory (IRT) method, to overcome such limitations, both for developing new tools, but also, to evaluate and improve existing tools. Rasch analyses give further insight into the character of the data and convert the ordinal data to linear measures. The basic assertions of the Rasch model are that the easier the item, the more likely it is to be passed by all persons, and the more able the person, the more likely he or she is to pass more difficult items. The use of Rasch measurement methods can ensure that the tests we are constructing are valid and reliable measures – unidimensional scales expressed in equal units of ability. When developing new or evaluating existing scales, a clear conceptualization of what one intends to measure and a careful evaluation of the test items is required. Do they evaluate what one really wants to measure and does each item represent a point along the same unidimensional continuum?

As pointed out by *Gustaaf Lankhorst* from Amsterdam, developing new instruments is a long and complicated job. Careful selection of existing instruments would therefore be the most obvious practice. However, a new generation of instruments might have attractive clinometric properties. There should also be convincing evidence that the new instrument is better, and that it will be easy to use and that cost-benefit ratio of the new measure is favourable, even if training/certification may become more common.

As an example of development of a new instrument *Massimo Penta* from Brussels demonstrated the use of Rasch methodology for creating a new scale for measurement of manual ability. An item pool was evaluated by experts and tested experimentally using Rasch analysis. Item with disordered thresholds and those, which were misfitted were excluded. In patients after stroke it was demonstrated that bimanual activities were better targeted to the ability of the patients than were unimanual activities (3). The importance that item hierarchy is invariant across patient subgroups was stressed.

Continuing to build upon these themes, *Alan Tennant* from Leeds discussed issues related to translation and the use of outcome measures in different countries. In most cases, tools are translated and applied without examining the validity of the translations. Tennant stressed the importance of using Rasch measurement methods to examine differential item functioning to ensure that test items work in the same way despite the group assessed (4).

The innovative use of computer-adapted testing (5) to implement psychometrically sound, but cost-effective, assessment was introduced by *Alan Jette* from Boston. Stressing the point that most current outcome instruments have floor and

ceiling effects and lack the required reliability and sensitivity, Jette proposed using IRT methods to develop core items banks and then evaluate each person using customized subset of items. The basic principle is that each person answers only questions that are appropriately targeted. If the first question is too easy, a somewhat more difficult one is presented. If that is too difficult, then a slightly easier one is presented. Continuing in this manner, the person needs to answer only enough questions to meet the desired level of accuracy.

*Linda Tickle-Degnan*, also from Boston, discussed practical applications of evidence-based practice, including evidence that an available outcome measure is appropriate for evaluating an individual client. An important point of her presentation was the idea that we often over-focus on group-oriented population means rather than recognizing the importance of individual variability. There is a need, she argued, for research that will enable us better to understand systematic variations among individuals in context.

Transitioning to methods that can be used to evaluate outcomes related to patient experiences and perceptions, *Ann Öhman* from Umeå provided an overview of qualitative methodology for rehabilitation outcomes. Qualitative research is concerned with how the social world is understood, interpreted and experienced. Öhman presented an example where the research focused on describing and analysing how persons with pain experienced a rehabilitation program for musculoskeletal pain (6).

Patient-directed outcomes can be defined as constructed by the patients within a framework laid out by the professionals. Between these and professional directed outcomes are measures which are constructed by professionals for self-completion by patients. As pointed out by *Alan Tennant* there are principal problems with patient-generated instruments, correct completion may be low, and they may be affected by educational levels and socio-economic factors. There are still a number of problems concerning their reliability and validity, the treatment of the assessments scales and the lack of evidence to support the validity of the arithmetic operations, which underpin such measures.

*Björn Gerdle* from Linköping reported on special aspects of outcome measures in chronic pain trials and listed core domains to be considered: pain, physical functioning, emotional functioning, patient ratings of improvement and satisfaction with treatment, symptoms and adverse effects, and patient disposition. Several outcomes are necessary, which should be validated

and well known. Subjective as well as objective variables should be used.

The seemingly easy way of dichotomizing outcome with “on sick-leave” or “back to work”, has its problems as pointed out by *Jan Ekholm* from Stockholm. This is partly due to the different regulations that govern the registration, for example how return to work part-time or vocational rehabilitation is classified. In addition, aspects of politics on unemployment statistics and sick-leave were addressed, which in the European societies might have an influence on each other.

The symposium ended with a panel discussion in which the invited speakers gave their views on what outcome analysis would be focussing on in 10 years time. Both the possibilities of using the ICF more, and combining qualitative and quantitative aspects were discussed, as well as the development of new scales and the need to assess participation and integration in society. There is also a need for research into the effects of contextual factors. Instruments ought to be based on modern psychometric theories and have a conceptual background. Outcome measures should be part of the education in different health professions. It was also pointed out that qualitative methods have a value in themselves in addition to being useful as a basis for quantitative questionnaires.

## REFERENCES

1. Stucki G, Grimby G, eds. ICF core sets for chronic conditions. *J Rehabil Med* 2004; (suppl 44): 1–141.
2. Cieza A, Brockow T, Ewert T, Amman E, Kollertis B, Chatterji S, et al. Linking health-related measurements to the International Classification of Functioning, Disability and Health. *J Rehabil Med* 2002; 34: 205–210.
3. Penta M, Tesio L, Arnould C, Zancan A, Thonnard JL. The ABILHAND questionnaire as a measure of manual ability in chronic stroke patients: Rasch-based validation and relationship to upper limb impairment. *Stroke* 2001; 32: 1627–1634.
4. Tennant A, Penta M, Tesio L, Grimby G, Thonnard J-L, Slade A, et al. Assessing and adjusting for cross cultural validity of impairment and activity limitation scales through Differential Item Functioning within the framework of the Rasch model: the PRO-ESOR project. *Med Care* 2004; 42 (suppl): 125–136.
5. Haley SM, Coster WJ, Andres PL, Kosinski M, Li P. Score comparability of short forms and computerized adaptive testing: Simulation study with the Activity measure for post-acute care. *Arch Phys Med Rehabil* 2004; 85: 661–666.
6. Gustafsson M, Ekholm J, Öhman A. From shame to respect: musculoskeletal pain patients' experience of a rehabilitation program, a qualitative study. *J Rehabil Med* 2004; 36: 97–103.

*Gunnar Grimby, Jan Ekholm, Anne Fisher and Katharina Stibrant  
Sunnerhagen  
Göteborg, Stockholm and Umeå, Sweden*