A comparison of measures of disability and health status in people with physical disabilities undergoing vocational rehabilitation

S. Kelly and E. G. Jessop

Abstract

Background The aim of the study was to test among people undergoing vocational rehabilitation six measures of disability and health status commonly used in medical rehabilitation.

Method A cross-sectional survey was carried out on 30 people with disabilities on an occupational programme at a non-medical rehabilitation facility in England. Measures used were the Barthel index, Extended Activity of Daily Living (EADL) scale, the Office of Population Censuses and Surveys (OPCS) disability questionnaire, the Functional Independence Measure (FIM), the Nottingham Health Profile (NHP) and the Medical Outcome Study Short Form 36 (SF36). The main outcome measures were item non-response and time for completion of each measure, profile of disability and health status described by each measure, including pertinent domains detected or missed, and floor and ceiling effects.

Results Item non-response was very low with all the instruments; the Barthel index was on average the quickest to complete (mean time 2.2 minutes) and the SF36 the longest (mean time 9.1 minutes). The study group were characterized as having problems in mobility or locomotion and bladder or bowel control, but some of the instruments were insensitive, detecting no disability in many subjects (e.g. 33 per cent showed no disability on the Barthel index). The SF36 scores were the least affected by floor and ceiling effects; mean SF36 scores on all scales except physical functioning were similar to those for the general population of similar age. Some problems were detected by only one of the instruments (e.g. pain and sleep by the NHP, problems with intellectual functioning by the OPCS scale).

Conclusion Disability measures commonly used in medical rehabilitation, such as the Barthel score and FIM, may be less useful in vocational rehabilitation where disabilities are less severe. Other measures show more promise but further testing is needed.

Keywords: disability, measurement, SF36

Background

Measurement of physical disability presents many problems, but such measurement is important not only clinically, to document progress in individuals, but also, and perhaps more importantly for public health workers, to allow assessment of how rehabilitation services are performing.

There are a considerable number of measures of health and rehabilitation status available, and some are routinely used in health services; for example, the Barthel index is widely used in the National Health Service, and the Functional Independence Measure in medical rehabilitation units in the United States.

Measures which are useful in one setting may, however, be less so in others; for example, Parker et al. found that although the Barthel index was a useful measure of outcome on a geriatric in-patient unit, it was insensitive to clinical change in the rather more able population of attenders at a day hospital for the elderly. We were interested to test a number of measures in common use in medical rehabilitation in the rather different setting of a vocational rehabilitation unit. The distinction between medical and vocational rehabilitation is to some extent artificial, but in general the primary focus of medical rehabilitation is medical aspects of physical disability, which can be characterized as treating impairment and reducing

Wessex Institute of Public Health Medicine, Romsey Road, Winchester SO22 5DR.

S. KELLY, Senior House Officer

E. G. JESSOP, Consultant

Address correspondence to Dr E. G. Jessop, at: West Surrey Health Commission, Ridgewood Centre, Old Bisley Road, Camberley GU16 5QE.

© Oxford University Press 1996
disability, whereas vocational rehabilitation focuses on the vocational skills needed to overcome handicap in return to work. Nevertheless, for funding organizations such as health authorities in the National Health Service (NHS) there is a clear administrative difference between medical rehabilitation services, usually located in hospitals, and vocational rehabilitation units run by social service departments or independent organizations. Aitken and Cornes have discussed the problems caused by this separation, but it is a fact of life. There would be obvious advantages if one instrument proved capable of measuring recovery from medical rehabilitation through vocational rehabilitation to return to work, though the different aims of medical and vocational rehabilitation may make this an unrealistic proposition. In particular, measures in common use in medical rehabilitation are likely to be directed at physical function whereas the whole point of vocational rehabilitation may be to make functional limitation irrelevant to work ability: to ensure that disability does not cause handicap.

Our study was designed to compare the use of different measures of disability in characterizing a group of people receiving vocational assessment and rehabilitation. Our selection of six measures was dictated by knowledge of a concurrent multi-centre trial of measures in medical rehabilitation (D. L. McLellan, personal communication, 1993), which we thought our study could usefully complement. McDowell and Newell have given a useful commentary on the user's perspective in evaluating a health measurement. We were interested in feasibility of administration and use, acceptability to respondents, response rates, and, as far as was possible in a cross-sectional study, sensitivity to change. We did not have the resources for a study large enough to evaluate metric properties formally, nor aspects such as construct validity and test–retest reliability.

Instruments and measures

Each study subject was scored on the following scales or instruments:

1. Barthel index: Activities of Daily Living (ADL);
2. Extended ADL;
3. Office of Population Censuses and Surveys (OPCS) disability measure;
4. Functional Independence Measure (FIM);
5. Nottingham Health Profile (NHP);
6. Medical Outcomes Study Short Form 36 (SF36).

The instruments were applied by an interviewer and completed by the person with a physical disability, assisted where necessary by their key worker. For the NHP and SF36 the items were asked verbatim; for the other instruments the information required was obtained by appropriate questioning of the subject. Thus all six measures were scored solely on information provided by, or according to the opinion of, the subject.

Data analysis

Barthel, NHP and SF36 scores were calculated in accordance with the appropriate instructions or manuals. The primary quantitative comparisons were of time to complete, item completion rate, and floor or ceiling effects, the last being defined as items for which 75 per cent or more of responses were at the scale minimum or maximum. Where two scales had comparable items (e.g. pain scales) Pearson correlation coefficients were calculated. Data were analysed with the Statistica™ software package.

Ethical consent

Ethical consent was obtained for the study from the Winchester Ethics Committee.

Results

Results were obtained from 30 subjects: 16 males and 14 females, with a mean age of 34.7 years (range 20–67 years). Ten eligible subjects were unwilling to be interviewed (response rate 30/40=75 per cent); the non-participants were 6 males and 4 females, with a mean age of 29.8 years (range 17–53 years).

Results by questionnaire, comparison against norms

Basic comparisons are shown in Table 1. The Barthel ADL index took between 1 and 6 minutes to complete (mean 2.2 minutes). There was no item non-completion. The distribution of scores was highly skewed, with 75 per cent or more of subjects...
TABLE 1 Comparison of measures

<table>
<thead>
<tr>
<th></th>
<th>Barthel</th>
<th>EADL</th>
<th>OPCS</th>
<th>FIM</th>
<th>SF36</th>
<th>NHP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (and range)</td>
<td>2-2</td>
<td>4-3</td>
<td>7-3</td>
<td>6-1</td>
<td>9-1</td>
<td>6-1</td>
</tr>
<tr>
<td>completion time in</td>
<td>(1-6)</td>
<td>(1-9)</td>
<td>(2-20)</td>
<td>(2-13)</td>
<td>(3-22)</td>
<td>(3-13)</td>
</tr>
<tr>
<td>minutes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item completion rate</td>
<td>100</td>
<td>89</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>96</td>
</tr>
<tr>
<td>(%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraction (and %) of</td>
<td>7/10</td>
<td>8/22</td>
<td>6/13</td>
<td>11/18</td>
<td>1/8</td>
<td>0/6</td>
</tr>
<tr>
<td>domains in scale</td>
<td>(70)</td>
<td>(36)</td>
<td>(46)</td>
<td>(61)</td>
<td>(12)</td>
<td>(0)</td>
</tr>
<tr>
<td>showing floor or</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ceiling effect*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domain coverage (see</td>
<td>Locomotion/mobility</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>text for details):</td>
<td>Eating</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Personal care</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bowel and bladder control</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pain</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cognition and consciousness</td>
<td></td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Emotional state and mental</td>
<td></td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>health</td>
<td></td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Behaviour and social functioning</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>Sleep</td>
<td></td>
<td></td>
<td></td>
<td>Y</td>
<td></td>
</tr>
</tbody>
</table>

*Defined as 75 per cent or more of responses at scale minimum (floor) or maximum (ceiling).

achieving the maximum score (i.e. being independent) on all items except continence of bowels and bladder and climbing stairs, and 50 per cent or more achieving independence on these three items. Eleven (33 per cent) subjects showed no disability on the ADL score.

The Extended ADL index took between 1 and 9 minutes to complete (mean 4-3 minutes). Responses of 'no opportunity' or 'not relevant' were obtained on 22 items, and classifying these as item non-completion, the overall item non-completion rate was 74/660 (11 per cent). The items for taking a drink from one room to another and managing one's own garden had the highest proportion of 'not relevant' responses.

The OPCS scales took between 2 and 20 minutes to complete (mean 7-3 minutes). There was no item non-completion, but administering the item for behaviour (requiring the respondent to admit, for example, to ‘outbursts of temper at other people with very little cause’) was unsatisfactory. Seventy-five per cent or more of subjects showed no disability on the scales for
judged by correlation between item scores and scale on Part II of the questionnaire. The item on sex life in NHP surveyed a general population sample and Part II of the questionnaire was dropped after adverse cent) of the Part I scores, and three item non-responses mental health scales had a high loading on one factor, poor?'), which was inversely correlated with the overall physical role function, vitality, social functioning and of the variance of each scale. In this sample, the scales. A sample size of 30 is adequate to detect a difference of 10-20 points on each scale with 80 per cent, scales. A sample size of 30 is adequate to detect a difference of 10-20 points on each scale with 80 per cent, power testing for significance at the 5 per cent level.

The SF36 was internally consistent in this sample, as judged by correlation between item scores and scale scores: with one exception, each item correlated more highly with the scale of which it forms a part than with any other SF36 scale. The anomaly was the first item ('is your health: excellent, very good, good, fair, poor?'), which was inversely correlated with the overall score for general health. Factor analysis on this small sample revealed an unusual factor structure to the results: in the general population the SF36 has two factors (labelled by the developers ‘physical health’ and ‘general health’) which account for a large proportion of the variance of each scale. In this sample, the physical role function, vitality, social functioning and mental health scales had a high loading on one factor, and only general health had a high loading on the other factor.

The Nottingham Health Profile took between 3 and 13 minutes to complete (mean 6.1 minutes). There was one instance of item non-response, though administering the questionnaire proved difficult. The distribution of responses was highly skewed, with 75 per cent or more of subjects achieving the maximum score on 13 of the 18 items, and 50 per cent or more achieving the maximum on four of the remaining five items. The lowest mean scores (indicating higher disability) were for use of stairs and bladder management.

The SF36 took between 3 and 22 minutes to complete (mean 9.1 minutes), and item non-response occurred on only three occasions. The distribution of scores was skewed, but not extremely so, except for the emotional role scale. The lowest mean score was for the physical function scale. Compared with published data from a general population sample aged 35–44 years, our sample scores markedly lower on the physical function scale, but is not significantly different on the other scales. A sample size of 30 is adequate to detect a difference of 10–20 points on each scale with 80 per cent, power testing for significance at the 5 per cent level.

The SF36 was internally consistent in this sample, as judged by correlation between item scores and scale scores: with one exception, each item correlated more highly with the scale of which it forms a part than with any other SF36 scale. The anomaly was the first item ('is your health: excellent, very good, good, fair, poor?'), which was inversely correlated with the overall score for general health. Factor analysis on this small sample revealed an unusual factor structure to the results: in the general population the SF36 has two factors (labelled by the developers ‘physical health’ and ‘general health’) which account for a large proportion of the variance of each scale. In this sample, the physical role function, vitality, social functioning and mental health scales had a high loading on one factor, and only general health had a high loading on the other factor.

The Nottingham Health Profile took between 3 and 13 minutes to complete (mean 6.1 minutes). There was some item non-response on both parts of the questionnaire, resulting in inability to calculate 8/180 (4 per cent) of the Part I scores, and three item non-responses on Part II of the questionnaire. The item on sex life in Part II of the questionnaire was dropped after adverse reaction to it in the pilot study. The developers of the NHP surveyed a general population sample and published scores in five-year age bands. Compared with people aged 35–39 years old, our sample has higher (worse) scores for pain, sleep, social isolation and physical mobility.

Comparison of questionnaires
The six instruments evaluated vary in coverage and depth. For some items the proportion of subjects who scored as having no disability or as being independent was so high that formal correlation tests have no value. For simplicity, results are presented under nine functional dimensions. These groupings are arbitrary: formal factor or cluster analysis to identify domains seemed inappropriate with our small sample size.

Locomotion and mobility
As noted above, the most severe disabilities detected by the scales were physical, particularly in mobility. The correlations between the ADL item on walking, the EADL mobility score, the OPCS and FIM locomotion item and the SF36 and NHP scales for physical function and mobility were not very high, except as follows (judged by a correlation coefficient greater than 0.75): (1) EADL mobility score correlates well with OPCS locomotion item and the FIM locomotion items; (2) the OPCS locomotion score correlates well with the FIM score for use of stairs, but not so well with the score for walking or chair use; (3) the FIM score for stairs correlates well with the NHP mobility score.

Eating
The ADL, EADL, OPCS and FIM have items for feeding; no subject reported any disability in this area on any scale, except for one positive response to the OPCS item.

Personal care
The ADL items for grooming and dressing respectively recorded dependence in one subject and a need for help in three; 27/30 scored no disability on the OPCS personal care item; and 26 or more were scored as independent on the FIM grooming and dressing items.

Control of bowel and bladder
The ADL, OPCS and FIM have items related to this; the scores on these scales were highly correlated. All subjects were independent in use of the toilet as judged by ADL and FIM; for transfers to WC, modified independence was recorded in five subjects, and supervision or set-up required in one subject, on the FIM item, and one subject recorded a need for major help on the ADL transfer item.
The four functional instruments (ADL, EADL, FIM and OPCS) do not measure pain. The NHP and SF36 both have pain scores, but these were poorly correlated in this sample (Spearman \( R \) correlation coefficient \(-0.3, p = 0.18\)). Unlike the SF36, the mean NHP score was higher than that for the general population, but this was due to a few high scores in a skewed distribution.

**Cognition**

ADL and EADL have no items for cognitive state. The OPCS scale has a score for consciousness (fits) which is not included explicitly on any of the other scales, but was a problem for eight subjects. The OPCS scores for intellectual functioning and communication are similar to the FIM items for problem solving and memory. The OPCS communication scale has a single item for communication, which is given two scores (comprehension and expression) in the FIM.

**Emotional state and mental health**

The NHP score for emotional reactions was poorly correlated with the SF36 mental health score (Spearman \( R \) coefficient \(-0.37, p = 0.05\)) and with the SF36 emotional role performance score (coefficient \(-0.05, p = 0.79\)), but as noted above our sample did not differ from the general population on these scores.

**Behaviour and social functioning**

ADL has no items in this category, but EADL has six questions in a category designated as 'leisure', and OPCS has scores for behaviour and communication. These items all correlated poorly with each other and with the NHP and SF36 social scores. The NHP score for social isolation was poorly correlated with its equivalent in the SF36 (social functioning): correlation coefficient \(-0.16, p = 0.39\). As noted above, this sample scored worse than the general population on the NHP but not on the SF36 for social isolation and functioning. Almost all subjects scored no problem on the FIM social interaction scale.

**Sleep**

Only the NHP has a sleep scale, and as noted above the sleep score in this sample was worse than that for the general population.

**Discussion**

This study compared the use of four measures of functional ability or dependence and two quality of life questionnaires, all in common use in medical settings, in a group of people undergoing vocational rehabilitation.

All of the instruments were useable, with low item non-response. The Barthel index was on average the quickest to complete (mean time 2.2 minutes) and the SF36 the longest (mean time 9.1 minutes). From the researchers' point of view, the FIM and OPCS instruments were the most difficult to use in this setting: all items in the other four instruments are (or can be phrased as) simple questions, whereas the FIM and, to a lesser extent, the OPCS instrument are designed for assigning a score on the basis of a set of information gathered.

The study group were characterized as having problems in mobility or locomotion and bladder or bowel control, but some of the instruments were insensitive, detecting no disability in many subjects (e.g. 33 per cent showed no disability on the Barthel index). The SF36 scores were the least skewed by floor and ceiling effects; mean SF36 scores on all scales except physical functioning were similar to those for the general population of similar age. Some problems were detected by only one of the instruments (e.g. pain and sleep by the NHP, problems with intellectual functioning by the OPCS scale).

The SF36 is one of the best documented questionnaires available. We found that in this group of people the internal consistency of the questionnaire remained high, but the factor structure was different from that for datasets obtained from the general population. Not too much can be made of this finding given the small sample size of our study, but it raises the question of whether the concept (or more precisely the construct) of health is different in people with physical disability, general health being more important than separate constructs of physical and mental health. Brazier et al. found good correlation between SF36 and NHP scores for pain and for mental (SF36) or emotional (NHP) health (correlation coefficients were \(-0.55\) and \(-0.67\), respectively); in our study there was no correlation between these measures, but it should be noted that the distribution of NHP scores is very skewed, which makes correlation in a small sample difficult to assess.

The study has some limitations. The size of the sample (30 respondents) limits statistical power. We were not able to carry out repeated interviews of the same subjects to assess test–retest characteristics of the questionnaire, and their sensitivity to change over time in health status of the subjects, which is obviously important for a rehabilitation service. Non-response is a concern in any survey, though the 75 per cent response rate we achieved is probably good enough for conclusions about the performance of these questionnaires in this group of people.

Most studies of disability reported in the medical literature rely on professional assessment of abilities: we used self-report, mostly because it seemed to us...
paternalistic to seek information from staff rather than question the person with a disability directly. There will undoubtedly be discrepancies between self report of disability and professional assessment; it is a matter of opinion which is more relevant." We noted a number of subjects with very abnormal gait who to us seemed to have great difficulty walking but who reported no difficulty in this respect on the questionnaires. A high proportion of the study population were regarded clinically as having some degree of intellectual impairment (R. G. S. Platts, personal communication, 1994), which may well have affected individuals' perceptions of themselves. At the very least, no direct comparison can be made between studies using self report and those using professional assessment of disability.

In sum, we have found that some measures commonly used in medical rehabilitation, such as the Barthel score and FIM, may be less useful in vocational rehabilitation where disabilities are less severe. Other measures, particularly the SF36, show more promise but failed to detect some problems such as those related to sleep. This may to some extent reflect a difference between the functional measures and those which tap quality of life. As noted above, function is a primary concern in medical rehabilitation, whereas it should in one sense be irrelevant to vocational rehabilitation.

We studied only one rehabilitation facility, and we think that this is a more important limitation to our main findings than the small sample size. Different facilities may serve very different types of patient or client. We would be keen to repeat the survey to see how well the instruments detect change over time.

Further research should focus on seeing whether our results apply to other facilities, and assessing the usefulness of the different instruments in monitoring change over time, both in individual patients and, through repeat censuses, in the characteristics of patients in a particular facility.

Acknowledgements

We are grateful to Dr Robin Platts, Medical Director of the vocational rehabilitation centre, and to the subjects who agreed to help us with this study. We also acknowledge gratefully the support and advice of Professor Lindsey McLellan, and the perceptive and helpful comments of two anonymous referees.

References

19 Skruppy M. Activities of daily living evaluations: is there a difference in what the patient reports and what is observed? Phys Occup Therap Geriat 1993; 11: 13–25.

Accepted on 5 December 1995