

ORIGINAL REPORT

## DISABILITY ASSESSMENT INTERVIEW: THE ROLE OF DETAILED INFORMATION ON FUNCTIONING IN ADDITION TO MEDICAL HISTORY-TAKING

Jerry Spanjer, MD<sup>1,2</sup>, Boudien Krol, PhD<sup>2</sup>, Roel Popping, PhD<sup>3</sup>, Johan W. Groothoff, PhD<sup>2</sup> and Sandra Brouwer, PhD<sup>2</sup>

From the <sup>1</sup>Social Security Office, Assen, <sup>2</sup>Department of Health Sciences, Graduate School for Health Research, University Medical Center Groningen and <sup>3</sup>Department of Sociology, Interuniversity Center for Social Science Theory and Methodology, University of Groningen, Groningen, The Netherlands

**Objective:** To investigate whether the provision of detailed information on participation and activity limitations, compared with medical information alone, influences the assessment of work limitations by physicians.

**Methods:** Three groups each of 9 insurance physicians used written interview reports to assess work limitations in 30 patients with low back pain or lower extremity problems. Each group was given different kinds of information on the patient: the first group received only medical information; the second group received detailed information on participation and activity limitations; and the third group was provided with both types of information. Agreement percentages within the groups and differences between the groups in scores given on the work limitation items of the Functional Ability List were measured.

**Results:** The groups showed no important differences in agreement percentages (mean percentage approximately 80%). The physicians who received either medical information or both forms of information indicated fewer work limitations compared with physicians using detailed information on participation and activity limitations.

**Conclusion:** Information on participation and activity limitations provided by the patient has only limited influence on inter-rater reliability. However, there was a significant difference in scores on assessed work limitation items compared with medical history-taking alone. Therefore, in disability assessment interviews physicians should ask for medical information as well as detailed information on participation and activity limitations.

**Key words:** disability evaluation, observer variation, outcome assessment, interview, reproducibility of results, medical history-taking.

J Rehabil Med 2009; 41: 267–272

Correspondence address: Jerry Spanjer, Ter Borch 34, NL-9472 RB Zuidlaren, The Netherlands. E-mail: jerry.spanjer@uwv.nl

Submitted May 22, 2008; accepted October 24, 2008

### INTRODUCTION

The assessment of a patient's work limitations is a complex task. Common ways to assess work limitations include self-

reports, medical interviews and examinations and functional testing methods. In all these methods of assessment validity and reliability are questionable where a disability benefit is concerned (1–4). There are indications that patients do not always assess their own work limitations objectively (5). Furthermore, if the patient has a financial interest they might not always be motivated to give their best performance (6), while diagnoses or medical findings alone are not sufficient to assess work limitations (7, 8).

In the Netherlands, an employer has to pay wages for 2 years if an employee is unable to work due to disability. After these 2 years the patient can apply for a social disability benefit. The disability benefit procedure begins with an assessment of the patient's work limitations by an insurance physician, who interviews the patient and performs a physical examination. In addition, information provided by the occupational physician who treated the patient during the first 2 years of disability, and information from the treating physicians, is often available (9).

The assessed work limitations are registered in a standardized list, the Functional Ability List (FAL) (10). The assessment of work limitations is significantly based on an interview with the patient (11). In the interview the insurance physician enquires about, among other things, medical history, specific complaints and problems in functioning. Previous studies indicate that there is considerable inter-doctor variation amongst insurance physicians in the assessment of work limitations based on an interview and physical and mental examination (12, 13). Physicians are trained to enquire about impairments and their aim is to determine a diagnosis. However, the diagnosis alone is not always an appropriate measure by which to assess work limitations and this is a possible source of variation in assessment between physicians.

In the International Classification of Functioning, Disability and Health (ICF) a distinction is made between impairments (problems in body function or structure as a significant deviation or loss), activity limitations (difficulties an individual may have in executing activities) and participation (involvement in a life situation) (14). Research in the Netherlands has shown that although insurance physicians have the opportunity to obtain detailed information on participation and activity limitations,

when interviewing the patient they only do so superficially (11). Thus, although the physicians should assess work limitations, during the interview they did not enquire thoroughly about the activity limitations experienced by the patient.

There is a possibility that inter-doctor variation in the assessment of work limitations is reduced when insurance physicians ask the patient in detail about activity limitations and participation. Moreover, the credibility of the patient's statements can then be assessed more easily. In juridical literature there are tools to assess the credibility of statements (15, 16). An important part of the analysis of statements is the Criteria-Based Content Analysis (CBCA) (17, 18). In the CBCA there are a number of criteria with which to assess the credibility of a statement, two important examples of which are the "logical consistency" and the "quantity of details" the patient presents during the interview. In occupational medicine there is also a pleading for inquiring after concrete and detailed examples of each patient's limitations in work and daily life, as a way to assess which, and to what extent, work limitations are present (19).

More insight into the value of concrete and detailed information on disability assessment may improve the reliability and validity of disability assessments in patients applying for a disability benefit. The aim of the present study is to investigate whether concrete and detailed information on participation and activity limitations, compared with medical information alone (both provided by the patient in an interview), influences inter-rater variability and the degree of assessed work limitations between physicians in disability assessment.

## METHODS

### Procedure

Three groups each of 9 Dutch social insurance physicians were asked to assess patients' work limitations and record them in the FAL (10) by presenting 30 written patient reports. All 27 physicians had to assess the same patients, but each group of 9 physicians received different sorts of information on the patients, i.e. only medical information, only information on functioning or both kinds of information. As a result a total of 810 patient assessments were obtained. Each physician was asked for the percentage of relevant information they thought was provided by each report (100% being all the information needed for a trustworthy assessment). Moreover, after each assessment the physicians were asked to indicate which specific information they thought was missing.

### Physicians

Out of a population of 524 Dutch social insurance physicians 30 were randomly sampled, stratified by region. Of these, 26 physicians were willing to co-operate, 3 were not able to co-operate due to long-term absence, and one did not feel motivated to participate. The physicians who declined participation were replaced by random sampling. Twenty-seven physicians returned a complete set of assessment lists; a response rate of 90%. The average length of time spent by these physicians in professional practice was 13 years (range 5–31 years).

### Patients

Thirty patients working in healthcare organizations or in retail and applying for a social disability benefit. The patients were randomly sampled, but only patients with low back pain or a lower extremity complaint were selected in order to obtain a homogeneous group while sufficiently completing the items of the FAL. Patients with these diagnoses represent approximately 30% of the entire population applying

for a social disability benefit. Half of the remaining population apply because of mental health problems and the other half have problems such as neck and upper extremity complaints, heart and lung diseases or cancer. The selected patients were diagnosed as follows: 12 with low back problems (spinal fracture, herniated disc, M. Scheuermann, chronic non-specific low back pain), 6 with fibromyalgia, 4 with knee problems, 2 with hip problems, 2 with rheumatoid arthritis and 4 with generalized arthrosis. The mean age of the patients was 48.9 years (range 30–63 years) and 80% were women. The mean duration of sick leave was 3.3 years (range 1–10 years). The patients had worked, on average, for 10.3 years in their last job (range 1–27 years) for 23.2 h a week (range 2–48 h a week).

### Reports

A written report comprised an interview with the disabled patient and a written report on physical examination. The interview was semi-structured and consisted of the following ICF items:

- *Impairments*: information on the patient concerning medical history, diagnosis, therapy and medication, progress of illness and medical complaints.
- *Activity limitations*: information on the patient concerning limitations experienced in daily life and work, such as, for example, lifting, walking and bending. The patient was asked for detailed and concrete examples of the limitations experienced.

An example: *Standing*. I can't stand very long. For example, I had to stand in line for concert tickets to see James Last. After 15 min my back ached and I had to step out of line, and my wife had to buy the tickets.

- *Participation*: information on the patient concerning activities of daily life (ADL), descriptions of a normal day, hobbies, housekeeping, social contacts and work. The patient was asked which activities were actually executed and for how long.

For instance: *Description of a normal day*. Yesterday I got up at 5.20 a.m., washed, got dressed and drank a cup of coffee. At 6.15 a.m. I went to work. Then I loaded the car with about 60 crates, each with eight loaves of bread. At 7.30 a.m. I drove off and went to 2 shops to deliver the bread. At 8.45 a.m. I came home and drank some coffee. For the remainder of the morning I did some housekeeping, which involved vacuum cleaning and mopping the floor. At 12 noon I walked the dog for about half an hour and had lunch. In the afternoon I read the newspaper, sat in the garden and read a book, drank tea and cooked dinner. At 6.00 p.m. I ate dinner and cleared the table. In the evening I watched Wimbledon on TV, walked the dog and went to bed at 10.30 p.m.

Three versions of reports were made for each of the 30 patients: a *medical* version with a summary of the interview regarding impairments and a description of the physical examination, a *functional* version with a summary of the interview regarding activity limitations and participation as well as the same description of the physical examination, and a *complete* version with all elements mentioned.

### Functional Ability List

The insurance physicians were asked to record their assessment of work limitations in the 36 physical items of the FAL. All insurance physicians were experienced at using the FAL. The items vary from a dichotomous scale to a 4-point scale. An example is the item "lifting or carrying":

#### Lifting or carrying

- 0: normal, can carry or lift about 15 kg (toddler)
- 1: slightly limited, can carry or lift about 10 kg (small toddler)
- 2: limited, can carry or lift about 5 kg (bag of potatoes)
- 3: severely limited, can carry or lift about 1 kg (1 l of milk)

### Analysis

The "linear weighted observed percentage agreement" on the FAL items was taken as a measure of inter-rater reliability between the

assessments of the insurance physicians. Due to the fact that the marginal distribution of the variables was highly skewed each time, the computation of an agreement index based on Cohen's kappa could not be used. A requirement for the use of this index is that the marginals have more or less the same frequency. If not, this will result in an overestimation of the expected agreement (20). The statistical software package AGREE 7.3 (21) was used for the calculation of the values. This package allows the calculation of an average "linear weighted percentage agreement" (22) between all pairs of raters. In general, an agreement percentage of 70% or higher is considered good, and greater than 90% is considered excellent (23).

The Mann-Whitney *U*-test was used for the between-group differences in height of scores on the items. This test investigates the difference in ordering of the assessments by the physicians in the different pairs of groups.

## RESULTS

Table I presents the average percentages of linear weighted agreement within the 3 groups of physicians as well as the significant differences in scores on the FAL items between the groups.

The group using the medical version had a mean percentage agreement of 80.1% (range 58–98%), the group using the functional version 81.3% (range 56–93%) and the group using the complete version 80.3% (range 57–95%).

In 11 out of the 21 items the physicians who were provided with the functional version gave more serious activity limitation scores in their assessments compared with the physicians who were either given the medical or the complete versions.

Significant differences were found between those who received the medical and the complete versions in 7 out of 21 items. Those using the medical version revealed more serious limitations 3 times, and the physicians using the complete version did so 4 times.

Table II presents the maximum amount of hours a patient can function in a day according to the physicians.

Within the group of physicians provided with the medical version, a limitation in the hours a patient can function daily was recorded 27 times (range 0–10 times/physician), within the group provided with the functional version 24 times (range 1–6 times/physician) and within the group provided with the complete version 6 times (range 0–3 times/physician).

When asked for the percentage of relevant information the physicians thought was provided by the reports, the physicians using the medical version, on average, indicated 71% (range per patient 57–86%, range per physician 58–94%). In the functional version the mean percentage of relevant information available was evaluated as 74% (range per patient 62–86%, range per physician 55–95%) and in the complete version it was evaluated as 84% (range per patient 70–95%, range per physician 74–91%).

Table III presents the information provided by the reports in each group and the additional information the physicians indicated they needed for their assessment in each of the 3 versions of the reports.

The physicians who were only provided with medical information mainly indicated a need for more information

Table I. "Linear weighted percentage agreement" between the physicians within the 3 versions of reports (columns 2–4) and significant differences between the 3 versions on scores of the Functional Ability List items (columns 5–7)

Items	Percentage agreement			Significant differences		
	Medical	Functional	Complete	F and M	F and C	M and C
<i>Body movement scale</i>						
Reaching	98	90	95	M↑		M↑
Frequent reaching	82	82	83	M↑	C↑	
Bending (degrees)	78	82	83		F↑	
Frequent bending	77	82	75	F↑	F↑	
Rotation*	78	90	74	F↑		C↑
Push or pull	74	81	75	F↑	F↑	
Lifting or carrying	84	84	78	F↑	F↑	
Frequent light lifting	79	87	83			
Frequent heavy lifting*	94	93	94			
Walking	85	81	86	F↑	F↑	
Sustained walking	85	86	87		F↑	
Climbing stairs	82	81	82	F↑	F↑	C↑
Climbing	87	81	84	F↑	F↑	
Kneeling*	78	83	87		F↑	M↑
<i>Body posture scale</i>						
Sitting	85	79	79	F↑		C↑
Prolonged sitting	79	75	78	F↑		C↑
Standing	83	81	81	F↑	F↑	
Prolonged standing	85	84	84	F↑	F↑	
Prolonged kneeling*	66	85	72			M↑
Prolonged bending*	58	56	57			
Working above shoulder*	66	63	70			
Mean	80.1	81.3	80.3			

\*Dichotomous data, other items are ordinal.

↑More serious limitations.

F: functional version; M: medical version; C: complete version.

Table II. Number of hours per week each patient could work in each of the 3 versions according to the physicians

Patient	Medical									Functional								Complete				
	Physician	1	2	3	4	5	6	7	8	9	10-12	13	14	15	16	17	18	19-24	25	26	27	
1-12																						
13		30							20			30	20	20		20						
14				20																		
15																30						30
16																	30					
17								20														
18																20						
19					30	30								30		30				30		30
20					30	30																
21					30								30									
22					30																	
23																						
24					30											10						
25				30	30	20	20	20	20	20		20	20	20		20					30	
26					20								30									
27					30																	
28		20	30			20		20	20	20				30	20		30				30	30
29													30		30		30					
30					30																	

Empty cell = full-time ( $\leq 40$ ).

concerning the patients' activities, disabilities experienced and a description of a normal day. The physicians who only used the functional information particularly indicated a need for additional information about therapy and medication. All 3 groups indicated a need for additional information from the treating physician in 21–34% of cases.

## DISCUSSION

A good inter-rater agreement on the items was found within all 3 groups and there were no clear differences in percentage agreement between the groups. However, there were significant differences in the item scores of the 3 groups. The physicians

provided with medical information either alone or in combination with functional information gave fewer work limitation scores than the physicians who received detailed information on participation and experienced activity limitations only.

The physicians who made their assessments based either on medical or functional information found that they had obtained approximately 71–74% of the total information needed. The physicians with only medical information found that they were lacking information on activities and disabilities experienced in two-thirds of the cases. The physicians with only functional information (and a diagnosis) were missing information concerning therapy, medication and information from the treating physician in one-third of the cases.

Table III. Information provided and percentage of additional information needed according to the insurance physicians ( $n = 3 \times 9$ ) in each of the 3 versions ( $n = 3 \times 30$ )

Items	Medical		Functional		Complete	
	Provided	Info need (%)	Provided	Info need (%)	Provided	Info need (%)
Medical complaints	+	4	–	8	+	5
Therapy and medication	+	1	–	37	+	1
Medical history	+	0	–	1	+	3
Course of illness	+	1	–	8	+	1
Problems in life/work	+	0	–	0	+	6
Activities in life/work	–	37	+	4	+	1
Description of normal day	–	66	+	0	+	3
Disabilities experienced	–	47	+	13	+	9
Work and reintegration	+	4	+	1	+	11
Patient's opinion	+	7	+	1	+	4
Physical examination	+	16	+	13	+	22
Observation	+	0	+	1	+	5
Info treating physician	–	19	–	31	–	27

Provided: Information is provided in patient report (– not present; + present).

Info need: Percentage of times the insurance physician mentioned additional information was needed in a total of 270 assessments for each version (9 physicians with 30 assessments each).

The physicians who assessed the complete versions of patient information found they had received about 84% of the total information they needed and particularly wanted extra information from the treating physician.

The fact that considerably fewer differences were found between the groups of physicians using the medical and complete versions compared with those using the functional version seems to indicate that medical information carries more weight than self-reported activity limitations. However, there are reasons why information on self-reported activity limitations and participation does play an important role in the assessment of work limitations. Firstly, the physicians indicated that they needed the information. In 66% of the assessments the physicians indicated a need for a description of daily activities. Secondly, for physicians who assessed the complete version, compared with the physicians having only medical information at their disposal, significantly different scores were found in one-third of the items. Furthermore, patient assessments based on the complete version, as opposed to the medical version, revealed a score on limitations in the amount of hours a patient can function a day more than 4 times less often (6 times vs 27 times).

In daily practice it is possible that different physicians collect different kinds of information in their interview with a patient. One physician may collect more medical information, while another collects more information on activity limitations. As seen in this study, the outcome of a disability assessment depends on the kind of information upon which the physicians base their assessment. Therefore, the satisfactory inter-rater reliability found in this study (within the groups of physicians that were provided with the same information) cannot be translated into daily practice. Inter-rater variability can be reduced if physicians collect the same information, by using a semi-structured interview, for instance.

An ongoing difficulty with the assessment of work limitations is the lack of a gold standard. Different assessment methods result in different outcomes. Performance tests and observations of performance result in fewer limitations compared with assessments based on medical information by physicians. In addition, self-report questionnaires result in the reporting of the most serious activity limitations (3–5). This is in line with the findings of this study: assessments based on self-reported activity limitations reveal more limitations than assessments based on medical information. To our knowledge no other literature is available on the use of subjective information from the patient on activity limitations and participation (made concrete by enquiring into detailed examples), alongside medical history, in disability assessment.

Each physician had to review 30 reports, which involved approximately 2 days work. In spite of this demanding task a response rate of 90% was obtained because they were released from their normal duties.

For practical reasons in this study the assessments were based on written reports. Insurance physicians in the Netherlands, however, interview and examine their patients themselves. That is why one has to be careful to interpret the results from this study into daily practice and why further

research is needed. Furthermore, due to the fact that only patients with lower extremity and low back complaints who had applied for a disability benefit were assessed, it would not be correct to assume that the results apply to other illnesses or to revalidation.

The assessing physicians in this study were provided only with information from the patient and a physical examination. In further studies it may be interesting to investigate the results achieved when combining information from an interview with the patient and information from the treating physician or performance tests.

In conclusion, we can say that information on participation and activity limitations in addition to medical information only has a limited influence on inter-rater reliability. However, insurance physicians who assessed activity limitations based on concrete and detailed information on participation and activity limitations in addition to medical information thought they had more relevant information and gave more serious limitation scores than physicians who only had medical information. Therefore, the combination of concrete self-reported limitations with medical information seems to be useful.

## REFERENCES

1. Grotle M, Brox JI, Vollestad NK. Functional status and disability questionnaires: what do they assess? A systematic review of back-specific outcome questionnaires. *Spine* 2005; 30: 130–140.
2. Pransky GS, Dempsey PG. Practical aspects of functional capacity evaluations. *J Occup Rehab* 2004; 14: 217–229.
3. Brokaw JP, Walker WC, Cifu DX, Gardner M. Sitting and standing tolerance in patients with chronic back pain: comparison between physician prediction and covert observation. *Arch Phys Med Rehab* 2004; 85: 837–839.
4. Brouwer S, Dijkstra PU, Stewart RE, Göeken LNH, Groothoff JW, Geertzen JHB. Comparing self-report, clinical examination and functional testing in the assessment of work-related limitations in patients with chronic low back pain. *Disabil Rehabil* 2005; 27: 999–1005.
5. Dowrick AS, Gabbe BJ, Williamson OD, Wolfe R, Cameron PA. A comparison of self-reported and independently observed disability in an orthopedic trauma population. *J Trauma* 2006; 61: 1447–1452.
6. Menard MR, Hoens AM. Objective evaluation of functional capacity: medical, occupational and legal settings. *J Orthop Sports Phys Ther* 1994; 19: 249–260.
7. Boot CRL, editor. Sick leave in asthma and COPD: the role of the disease, adaptation, work, psychosocial factors and knowledge. (dissertation) Nijmegen, The Netherlands: Radboud University Nijmegen; 2004.
8. Swinkels RA, Bouter LM, Oostendorp RA, van den Ende CH. Impairment measures in rheumatic disorders for rehabilitation medicine and allied healthcare: a systematic review. *Reumatol Int* 2005; 25: 501–512.
9. de Boer WEL, Besseling JJM, Willems JHBM. Organisation of disability evaluation in 15 countries. *Pratiques et Organisations de Soins* 2007; 38: 205–217.
10. ClaimBeoordelings – en BorgingsSysteem (CBBS) [Funcional Ability List]. Amsterdam: Landelijk Instituut Sociale Verzekeringen (LISV); 2002 (In Dutch).
11. Bont AA de, Berendsen L, Boonk MPA, Brink JC van den. In de spreekkamer van de verzekeringsarts. Een onderzoek naar het verzekeringsgeneeskundige deel van de WAO-claimbeoordeling. [In the consulting room of the insurance physician. A study of the role of

- the insurance physician in disability benefit]. Zoetermeer: College van Toezicht Sociale Verzekeringen (CTSV); 2000 (in Dutch).
12. Brouwer S, Dijkstra PU, Gerrits EHJ, Schellekens JMH, Groothoff JW, Geertzen JHB, Göeken LNH. Intra- en inter-beoordelaarsbetrouwbaarheid FIS-belastbaarheidspatroon en Functionele Mogelijkheden Lijst. [Inter- and intra-rater reliability of the 'Functional Information System' and the 'Functional Ability List']. *Tijdschrift voor Bedrijfs – en Verzekeringsgeneeskunde* 2002; 11: 360–367 (in Dutch).
  13. Spanjer J. De reproduceerbaarheid van WAO-beoordelingen, een literatuuronderzoek. [Reproducibility in disability assessment: a literature search]. *Tijdschrift voor Bedrijfs – en verzekeringsgeneeskunde* 2001; 9: 195–198 (in Dutch).
  14. World Health Organisation (WHO). International classification of functioning, disability and health. WHO: Geneva; 2002.
  15. Köhnken G. Statement validity analysis and the detection of the truth. In: Granhag P, Strömwall L, editors. *The detection of deception in forensic contexts*. New York: Cambridge University Press; 2004, p. 41–63.
  16. Gödert HW, Gamer M, Rill HG, Vossel G. Statement validity assessment: inter-rater reliability of criteria-based content analysis in the mock-crime paradigm. *Legal Criminol Psychol* 2005; 10: 225–245.
  17. Tredoux C, Pooley S. Credibility assessment. In: Tredoux C, Foster D, Allan A, Cohen A, Wassenaar D, editors. *Psychology and law*. Lansdowne, South Africa: Juta and Co. Ltd; 2005.
  18. Vrij A. Criteria-based content analysis: a qualitative review of the first 37 studies. *Psychol Public Policy Law* 2005; 11: 3–41.
  19. Spanjer J. Voorbeelden: een manier om de ernst van beperkingen te beoordelen. [Examples: a way to assess the seriousness of work limitations]. *Tijdschrift voor Bedrijfs – en Verzekeringsgeneeskunde* 2002; 10: 109–111. (Readers comments 2002; 10: 162–163) (in Dutch).
  20. Popping R. On agreement indices for nominal data. In: Saris WE, Gallhofer IN, editors. *Sociometric research, volume 1, data collection and scaling*. London: MacMillan Press; 1988, p. 90–105.
  21. Popping R. AGREE, a package for computing nominal scale agreement. *Comput Stat Data Anal* 1984; 2: 182–185.
  22. Cicchetti DV. A new measure of agreement between rank ordered variables. *American Statistical Association, Proceedings of the 80th annual convention. Proceedings of the annual convention of the American Psychological Association* 1972; 7: 17–18.
  23. Landis JR, Koch GG. The measurements of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174.