The disability assessment structured interview
Spanjer, Jerry

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2010

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Inter-rater reliability in disability assessment based on a semi-structured interview report

Jerry Spanjer
Boudien Krol
Sandra Brouwer
Johan W. Groothoff

Published in: Disability and Rehabilitation 2008;30:1885-90.
ABSTRACT

Objective
To investigate 1. if physicians find that they are able to perform a disability assessment based on written reports; 2. the inter-rater reliability between physicians in the assessment of work limitations.

Method
Twelve insurance physicians used written reports to assess work limitations in 12 patients. The reports concerned a semi-structured interview executed by a nurse practitioner. The insurance physicians were asked whether they could make reliable assessments based on these reports. In addition, inter-rater reliability was measured by computing their percentage agreement with respect to the mental and physical items of two Dutch disability lists (the Functional Information System and the Mental Ability List).

Results
The quality of the reports was evaluated as reasonable to good. Half of the physicians found the assessment based on the reports reasonably reliable, 25% found the opposite and 25% was indecisive. The overall agreement between the insurance physicians was reasonable to good with a mean agreement of 76% (range 64–88%). Agreement between the physicians concerning the number of hours a patient could function daily was low.

Conclusions
Half of the physicians thought that a reliable assessment based on the written information was possible. The quality of written patient reports made by nurse practitioners trained in conducting a semi-structured interview was considered reasonable to good by insurance physicians. The inter-rater reliability between insurance physicians of physical and mental disability assessment based on the written reports was reasonable to good. The assessment of the number hours patients could function daily had low inter-rater reliability.
INTRODUCTION

In the Netherlands an employer has to continue to pay an employee for two years if an employee is disabled from work. After two years, the patient can apply for social disability benefit. The procedure to assess disability benefit claims is as follows.

The insurance physician interviews and examines the patient to assess work limitations. The work limitations are recorded in a standardised list – the Functional Information System (FIS). The insurance physician registers in this list which work limitations the patient has and to what degree. For instance, if a carpenter has chronic lower back pain, the insurance physician has to evaluate the ‘Lifting’ item on a range from 1kg to more than 25kg.

Subsequently, a labour expert examines which jobs the patient is able to perform with the work limitations as assessed by the insurance physician. A computer matches the work limitations with a database of 7000 occupations which are described individually with the characteristics of each item being specified in detail. For example, the ‘Lifting’ item under the bridgekeeper occupation is described as requiring lifting up to 3kg regularly and up to 10kg occasionally.

The occupations selected by the computer are assessed by the labour expert on their suitability for the individual patient. The ultimate disability benefit will be the percentage ‘loss of ability to earn’, i.e. the difference between what the patient’s income was before his or her illness and what he or she is theoretically still able to earn. For instance, a carpenter who used to earn EUR 3000 a month and who is theoretically able to earn EUR 1500 as a bridgekeeper receives a 50% disability benefit.

The disability assessment by the insurance physician is based on an interview and an examination of the patient. Furthermore, the insurance physician can often obtain additional information from the treating physician/specialist and from the occupational physician who assessed the first 2 years of work disability.

The decision of the insurance physician is mainly based on a patient interview. During this interview, the insurance physician asks questions on medical impairments, the limitations the patient experiences and handicaps.

Three interview models are described in the Netherlands. The insurance physician often uses parts of the three different models in daily practice.

One of the models is the ‘Disability Assessment Structured Interview’ (DASI). This is a semi-structured interview in which the interviewer gathers information on the following items:

- occupation – the content of the occupation and how it is experienced by the patient
- impairments – medical history, the treating physician’s diagnosis, medication use, complaints and treatment
- activity limitations – limitations experienced in daily life and at work, e.g. lifting, walking and bending. The patient is asked for concrete and detailed examples of the limitations experienced.
• participation – activities of daily living (ADL), description of a usual day, hobbies, housekeeping, social contacts and work. The patient is asked which activities are actually performed and for how long
• the patient’s opinion about his work limitations.

In earlier research, 14 video recordings of DASI interviews were shown to 22 insurance physicians. The inter-rater reliability of the insurance physicians on the FIS items was reasonable to good (range 56–85%; mean 74%)⁵. A study where four insurance physicians assessed the work limitations of 30 chronic lower back pain patients showed a much lower agreement percentage on the FIS items (range 23–57%; mean 37%)⁶. In this latter study the insurance physicians performed the patient interviews themselves.

Research into other methods for assessing work-related limitations showed considerable differences in limitations between self-report, clinical examination and functional testing⁷. The issue of assessment of work limitations is an ongoing challenge, and in most countries the physician is the ultimate decision maker. Determining work limitations is complex because they cannot be measured by physical examination or be deducted from a diagnoses – there only is a modest relationship between disease and disability⁸-¹⁰. Often a wide inter-rater variability between physicians is present when assessing work limitations¹¹,¹².

A study in the Netherlands investigated the possibility of nurse practitioners taking over part of the insurance physicians’ tasks. The nurse practitioners were trained in interviewing patients according to the DASI method. Based on written reports of these interviews and on the physician’s own physical examination, the insurance physician assessed the work limitations.

The results of the study are described in this article in terms of the inter-rater reliability of disability assessment based on a written interview report according to the DASI method.

The research questions formulated are twofold:

1. Do insurance physicians find that they are able to perform a disability assessment based on the abovementioned written reports?
2. What is the inter-rater reliability of disability assessment between insurance physicians based on the abovementioned reports?

METHODS

Twelve insurance physicians received 12 reports from trained nurse practitioners assigned to assess each patient’s work limitations.

Patient selection

Four patient reports were randomly selected for each of the three participating nurse practitioners. Of these four reports, two patients presented mental complaints and two patients presented
physical complaints. As a result, 6 out of 12 patients were diagnosed as having mental problems (depression (twice), prolonged grief, migraine, social phobia and burnout) and 6 patients were diagnosed as having physical problems (lower back pain, shoulder, knee and feet complaints, RSI and breast cancer).

Reports
The reports consisted of an outline of the patient’s history regarding social security benefit, the report of the semi-structured interview, patient observations by the nurse practitioner and a physical examination by a physician.

Nurse practitioners
The nurse practitioners were trained in conducting the DASI interview during a five-day training exercise which covered instruction, attending physicians’ interviews, and the making and analysis of individual interview audiotapes. In addition the nurse practitioners were educated in guidelines, legal knowledge and conversation skills. After training was completed, continuous feedback on the interview reports was provided.

Raters
12 volunteer insurance physicians from different social security offices in the Netherlands with 3 to 27 years of experience (mean 12 years) in assessing work limitations. All 12 reports were sent to each of the 12 insurance physicians (response rate 100%).

Assessment
The insurance physicians were asked to record their assessment of work limitations in the Functional Information System (FIS). All insurance physicians were experienced in using the FIS. Since mental work limitations cannot be recorded in detail in the FIS, the Mental Ability List (MAL)\(^\text{13}\) was used to record mental work limitations. All 27 physical work limitations can be scored on a range from 2 to 10 in the FIS. For instance, the ‘Lifting’ item, which ranges from 1 to over 25kg. The 8 mental work limitations in the MAL can be scored on a range from 3 to 5.

After filling out these instruments, the physicians were asked to fill out an additional questionnaire in which their current experience in assessing work limitations based on the written report, quality of the reports and the perceived reliability of the assessment according to the physicians was recorded.

Analysis
The agreement percentages on the FIS and MAL items were calculated to assess inter-rater reliability between the 12 insurance physicians. The physical items percentage agreement was based on the 6 patients with physical complaints, whereas the MAL mental items percentage agreement was based
The statistical software package AGREE 7.0\textsuperscript{14-16} was used to compute percentage agreement. This statistical technique calculates a ‘weighted percentage agreement’ between multiple raters. Generally, a percentage agreement of less than 60\% is considered poor, 60 to 80\% is considered reasonable to good and more than 80\% is considered excellent\textsuperscript{17}.

**RESULTS**

*Questionnaire*

The quality of the reports was evaluated as reasonable to good by 11 physicians – 1 physician indicated reasonable to bad. Ten insurance physicians found that they had sufficient information to make assessments – 2 insurance physicians found that they had insufficient information in a number of reports. In some cases the physicians indicated that they needed more information – a more extensive description of the examination, physical or mental (5 times), information from the treating physician (3 times), a description of the illness before absence from work (3 times), and more detailed information about the patient’s occupation (2 times).

Each assessment took an average of 15 minutes (range 10–20 minutes). Half of the physicians found the assessment based on the reports reasonably reliable, 25\% found the opposite and 25\% were indecisive. Ten of 12 physicians found that a ‘live’ assessment clearly provided more information and two disagreed with this statement.

*Assessment*

In 2 of the 144 assessments (12 physicians, each with 12 reports) the physicians indicated that they did not have enough information for an assessment, in 3 the physicians found that the patient in question was not disabled. In all the other cases the physicians completed the FIS and the MAL. A patient with depression was considered fully incapacitated by 3 of the 12 physicians and a patient with a social phobia by 9 physicians.

Table 1 presents the agreement figures between the physicians on the most relevant items. The agreement percentages were reasonable to good with an average of 76\% (range 64–88\%). If an insurance physician found that a patient could only function a limited number of hours daily, he or she indicated a so-called ‘hours limitation’. The hours limitations as indicated by the physicians are presented in Table 2.

For 8 patients, 75\% of the physicians agreed that the patient was fully incapacitated or had an hours limitation. For 4 patients, half of the physicians agreed on an hours limitation, the other half found that the patient could function all day. They were ambiguous as to the number of hours to which the patient was restricted. Patients with mental illness were evaluated with an hours limitation or were considered fully incapacitated by an average of 4.7 physicians, whereas patients with physical limitations were evaluated as such by 1.3 physicians.
### Table 1. Percentage agreement between the physicians (N=12)

<table>
<thead>
<tr>
<th>Items</th>
<th>Percentage agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sitting</td>
<td>88</td>
</tr>
<tr>
<td>Standing</td>
<td>71</td>
</tr>
<tr>
<td>Walking</td>
<td>77</td>
</tr>
<tr>
<td>Climbing stairs</td>
<td>87</td>
</tr>
<tr>
<td>Climbing</td>
<td>69</td>
</tr>
<tr>
<td>Kneeling</td>
<td>67</td>
</tr>
<tr>
<td>Sustained bending</td>
<td>77</td>
</tr>
<tr>
<td>Frequent bending</td>
<td>76</td>
</tr>
<tr>
<td>Reaching</td>
<td>73</td>
</tr>
<tr>
<td>Working above shoulder</td>
<td>64</td>
</tr>
<tr>
<td>Lifting</td>
<td>80</td>
</tr>
<tr>
<td>Carrying</td>
<td>80</td>
</tr>
<tr>
<td>Structure</td>
<td>74</td>
</tr>
<tr>
<td>Responsibility</td>
<td>82</td>
</tr>
<tr>
<td>Time pressure</td>
<td>81</td>
</tr>
<tr>
<td>Emotional pressure</td>
<td>71</td>
</tr>
<tr>
<td>Concentration</td>
<td>81</td>
</tr>
<tr>
<td>Environment</td>
<td>72</td>
</tr>
<tr>
<td>Conflict handling</td>
<td>75</td>
</tr>
<tr>
<td>Social interaction</td>
<td>78</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td><strong>76</strong></td>
</tr>
</tbody>
</table>

### Table 2. Number of hours each patient could work according to the physician

<table>
<thead>
<tr>
<th>Physician</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5x4</td>
<td>5x4</td>
<td>5x4</td>
<td></td>
<td></td>
<td>5x6</td>
<td>5x6</td>
<td>5x4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>5x5</td>
<td></td>
<td>5x6</td>
<td>5x4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>5x6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5x4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>FI</td>
<td>5x4</td>
<td>5x4</td>
<td></td>
<td></td>
<td>FI</td>
<td>FI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>5x4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5x4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>FI</td>
<td>FI</td>
<td>FI</td>
<td>FI</td>
<td>FI</td>
<td>12</td>
<td>5x1</td>
<td>FI</td>
<td>FI</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>3x4</td>
<td>3x4</td>
<td>5x4</td>
<td>3x4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5x4</td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>5x4</td>
<td>5x4</td>
<td>5x4</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
</tbody>
</table>

N = Total number of patients. 65, 4, 7, 2, 1, 2, 2, 6, 3, 2, 4, 0

Fl = Fully incapacitated; 5x4 = 5 days x 4 hours; Empty cell = full time
DISCUSSION

The first question raised in this study was whether insurance physicians feel that disability assessment based on written reports can be performed reliably. Most of the insurance physicians involved found that the quality of the written reports was reasonable to good. Most of them also found that sufficient data was available to assess work limitations. However, they pointed out that a more extensive description of the physical and mental examinations and additional information from the treating physician would have been helpful. In other words, this indicates that the physicians wanted to check the external consistency of the patient’s story.

Though nearly all physicians found that ‘live’ assessment provides distinctly more information than a written report, 50% of the physicians indicated that an assessment based on written information can be performed reliably, 25% thought it was not very reliable and 25% were indecisive. Only in 2 of the 144 assessments did the physicians indicate insufficient data for assessment, which implies that the information provided was usually sufficient to assess work limitations.

The second question of this study concerned the inter-rater reliability between the insurance physicians of the disability assessment based on a written report. The agreement between the insurance physicians was reasonable to good with a mean of 76% for physical and mental work limitations. The results of the present study are comparable with findings in previous research where video recordings of DASI interviews were shown to insurance physicians. The mean percentage agreement in the present study was 76%, whereas the percentage agreement obtained in the video recordings study was 74%. However, a low inter-rater reliability was observed for the physician’s opinion on ‘the hours a patient can function daily’, not only as to whether an hours limitation was necessary as such, but also on the daily number of hours. This finding is in line with other research.

In daily practice this is often a point of debate between insurance physicians. Apparently, this problem is not only associated with performing an assessment based on written reports. Despite the ‘Dutch Guidelines for hours limitations’ for insurance physicians, there remains too much scope for subjective interpretation.

In this study we used the FIS and the MAL to record physical and mental work limitations. Due to the lack of a golden standard, we decided to use these methods, because both instruments are in use in Social Medicine in the Netherlands. They are developed after study of relevant literature, years of comments and feedback of insurance physicians and extensive use in daily practice. Furthermore, their content validity seemed sufficient, because they both describe the most important (physical and mental) demands which are relevant to daily work functioning and may be affected by physical or mental complaints. The FIS is based on the Dictionary of Occupational Titles taxonomy (DOT). In general, most physical work-related instruments are based on the DOT. This taxonomy has been described by the US Department of Labour and has gained support in many countries. Unfortunately, other information about the psychometric properties of these methods used in Social Medicine are still not available.
We decided to use the DASI method because the content validity of this method seemed sufficient while it is based on the levels of the International Classification of Functioning, Disability and Health (ICF). The ICF is a useful framework to understand the impact of a disease on patient’s health status. It describes health and health status in terms of functioning and disability. The items of the DASI are similar to the levels of disability described in the ICF - it involves dysfunction at one or more of the levels of impairments, activity limitations and participation restrictions influenced by environmental and personal factors. Based on the results of this research, it can be concluded that not only the content validity but also the inter-rater reliability is acceptable.

Because the participating physicians volunteered in the present study, selection bias may be possible. However, the physicians were working in different regions in the Netherlands and both experienced and less experienced physicians responded. Therefore, it is likely that the participating physicians are a reasonable reflection of the total population of physicians.

This study indicates that disability assessment based on a written report can be performed with an acceptable inter-rater reliability between insurance physicians. It should be borne in mind that in this study all participating physicians received the same written information. In the Netherlands in daily practice, insurance physicians interview patients themselves and the information they obtain can differ depending on the kind of questions asked or their individual conversational skills. In a study in which insurance physicians performed the interviews themselves without a structured interview, the mean percentage agreement was only 37%. This low percentage agreement can possibly be explained by the fact that the information on which the physicians based their assessments differed due to different interview styles. Whether training in performing semi-structured interviews can elevate the inter-rater reliability between insurance physicians is a subject for future research.

CONCLUSION

The quality of written patient reports made by nurse practitioners trained in conducting a semi-structured interview were considered reasonable to good by insurance physicians. Half of the physicians thought that a reliable assessment based on the written information was possible.

The assessment of work limitations by insurance physicians based on these semi-structured interview reports had a reasonable to good inter-rater reliability. This was the case with both physical and mental work limitations. Assessment of the number of hours a patient can function daily had a low inter-rater reliability.
Reference List


