The Disability Assessment Structured Interview

Its reliability and validity in work disability assessment

Printed by: Van Denderen, Groningen

Layout: P. van der Sijde, Groningen

Cover Photo: H. van Riezen

ISBN: 9789460700095

RIJKSUNIVERSITEIT GRONINGEN

The Disability Assessment Structured Interview

Its reliability and validity in work disability assessment

Proefschrift

ter verkrijging van het doctoraat in de
Medische Wetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. F. Zwarts,
in het openbaar te verdedigen op
woensdag 17 maart 2010
om 14.45 uur

door

Jerry Spanjer

geboren op 29 november 1958
te Groningen

Promotor:	Prof. dr. J. W. Groothoff
Copromotores:	Dr. B. Krol Dr. S. Brouwer
Beoordelingscommissie:	Prof. dr. M. H. W. Frings-Dresen Prof. dr. J. H. B. Geertzen Prof. dr. J. H. B. M. Willems

Paranimfen	Jan Buitenhuis Lex van de Ven

Contents

CHAPTER 1

Introduction

Work disability assessment in the Netherlands

In the Netherlands, an employer is obligated to continue to pay an employee for two years if an employee is disabled and unable to work. The employer is also responsible for reinstating the employee in his/her former work position or reintegrating that employee into another suitable work situation. An occupational physician assists the employer with medical questions regarding the employee's functional limitations and prognosis. After two years, the employee can apply for workers' compensation benefit from the Social Security Office. The benefit is based on the loss of wage-earning capacity by the employee. This is the difference between what the employee's income was before the sick leave, and what he or she is theoretically still able to earn in suitable work. The first step in this work disability assessment is an assessment of the patient's work limitations by an insurance physician. These work limitations are recorded in a standardized list – the Functional Ability List (FAL)[1]. In this list the insurance physician registers what work limitations the patient has and their extent. In the next step, a labor expert examines which jobs the employee is still able to perform despite the work limitations as assessed by the insurance physician. The labor expert is supported by a computer which matches the work limitations as listed in the FAL with a database of 7,000 occupations that describes the job demands in detail. The occupations selected by the computer are assessed by the labor expert as to their suitability for the individual employee.

The FAL is an important instrument in communicating between insurance physicians and labor experts. The FAL consists of a list of 70 different mental, physical and social items entailed in functioning on the job that are, in turn, grouped into 6 functional domains (Box 1). Each item can be rated as a nominal or ordinal variable on a two-to-ten scale. One example is the item "walking" found in the domain "dynamic movement", in which the insurance physician has to choose among four gradations (Box 2). In Addendum 1 the complete FAL can be found.

Box 1. Domains of the Functional Ability List (FAL)

I. Personal functioning (9 items)

II. Social functioning (12 items)

III. Adjusting to physical environment (10 items)

IV. Dynamic movement (24 items)

V. Static movement (11 items)

VI. Working hours (4 items)

Box 2. An example: the item "walking"

Walking

0 normal, can walk roughly one consecutive hour (a walk)

1 slightly limited, can walk for roughly 15-30 consecutive minutes (a stroll)

2 limited, can walk for roughly 5-15 consecutive minutes (to the mailbox)

3 very limited, can walk for less than 5 consecutive minutes (indoors)

Assessing work limitations

An insurance physician assesses the patient's work limitations. Several instruments for assessing functional work limitations are described[2], including functional capacity evaluations[3], self-assessment questionnaires[4;5] and assessments by physicians[6]. Each instrument has its own drawbacks. In functional capacity evaluations (standardized tests which measure actual physical performance) the sincerity of the patient's effort, the ability to perform work outside a laboratory setting, and whether the activities are considered medically safe can be questioned[7]. In self-assessment questionnaires, it is questionable whether patients are always being truthful, given the financial interest in the outcome. In physicians' assessments, reliability and validity are also questionable[8;9].

The determination of work limitations is complicated because symptoms or a diagnosis cannot simply be translated into functional limitations[10-14]. The one patient with rheumatoid arthritis, for instance, may have limitations that are quite different than the other's. Moreover, several diseases in which patients experience severe limitations have limited or no objective medical findings to back them up, for instance, chronic non-specific low back pain. Because of the medical knowledge needed to deal with these complicated aspects, in the Netherlands specialized insurance physicians assess the work limitations when workers' compensation is claimed.

Insurance physicians in the Netherlands base their assessments on certain information.

- a report from the employer in which the course of the first two years of work disability is summarized.
- medical information from the occupational physician who attended to the first two years of work disability.
- often, but not always, information from treating physicians (such as the primary care physician or specialists). This is only available if the occupational physician has requested this information (in about half of the cases). The insurance physician can always request medical information from the treating physicians if the patient agrees to this.
- every patient is seen by the insurance physician for an interview, observation and, in case of somatic complaints, a physical examination.

The decision of the insurance physician concerning work limitations is based, for the most part, on the patient interview[15;16].

The assessment interview

As part of the assessment interviews, insurance physicians use standard medical history-taking, including inquiring after symptoms, therapeutic interventions and medication. In addition, they specifically focus their attention on activity limitations and participation restrictions. For instance, the patient is asked how a normal day is spent.

Three defined interview models are described in the Netherlands[17], but the insurance physicians often use various parts of the three different models in daily practice[16]. We will give a short description of the three models:

- Methodical Assessment Interview[18]: The interview is semi-structured and has 10 topics including work possibilities, motivation, personal ideas about the pathology, vitality, personal changes, life events, thoughts about the future, medical history, work history and a description of a normal day. The arguments by the patient for the claim are important, with an emphasis placed on the functional limitations and abilities described in the claim. The patient is responsible for his own disability and recovery.
- Multi-causal Analysis[19]: This is an interview with a limited structure that includes five broad fields which can be interchanged. These fields include medical history and complaints, functioning, personal characteristics, work factors and personal factors. The physician engages the patient in the interview, and has an attitude of involvement, respect and attention. Perception and understanding of the patient are important.
- Disability Assessment Structured Interview (DASI)[20]: This is a semi-structured interview protocol with fixed topics which are largely based on the International Classification of Functioning Disability and Health (ICF)[21] (Addendum 2). The main topics are: introduction, work, impairments, the limitations to activity that are experienced, participation, the patient's opinion, and the physician's opinion. Each topic is subdivided into other topics. Concrete and detailed examples play important roles in defining the patient's limitations and abilities.

Reliability and validity

The assessment of functional limitations has major consequences. Therefore, it is imperative that different insurance physicians come to the same assessment (reliability), and that the proper functional limitations are assessed (validity). However, as far as the assessment of functional limitations by physicians in the Netherlands is concerned, no literature on reliability and validity can be found[17].

Reliability is the extent to which a test is able to measure in a consistent way, free from error. This

consistency may be either over time or between raters. There are several types of reliability including intra-rater reliability and inter-rater reliability. Intra-rater reliability examines the stability of data recorded by one person across two or more testing occurrences. Inter-rater reliability determines the variation between two or more raters who are assessing the same occurrence of the test[22;23].
Validity is the extent to which an instrument measures what it is intended to measure. There are several forms of validity including content, criterion (concurrent and predictive), and construct validity. Content validity is the degree to which test items represent the domain the test is intended to measure. This is usually determined by a panel of experts. Criterion validity is the extent to which the test performance is related to some other measure. It is comprised of concurrent and predictive validity. Concurrent validity examines the correlation between two or more measures given to the same subjects at the same time. Predictive validity compares a subject's performance on a test to performance at a future criterion. Construct validity is the extent to which a test can be shown to measure a hypothetical construct[23;24].

Aim of this thesis

The interview by the insurance physician plays an important role in the assessment of work disability, yet no studies into the reliability and validity of the interview have been conducted. The aim of this thesis is to study the psychometric qualities (reliability and validity) of the DASI method as part of work disability assessment in the Netherlands. Of the three interview models, we chose the DASI because it has detailed instructions (Addendum 2), and also because the author of this thesis is the developer of the DASI and is experienced in training insurance physicians in the DASI method. First, we wanted to describe a model that could identify sources of differences among physicians in their assessments of functional limitations, and then go on to systematically research the literature for instruments to assess functional limitations in workers' compensation claimants. This has resulted in the following research questions:

1. What are the possible sources of variation in work disability assessment?
2. Which instruments are described that measure or assess functional limitations in claimants, and what are their psychometric qualities?
3. What effect does detailed information on functioning in addition to medical history-taking have upon the functional limitations assessed and on inter-rater reliability?
4. In their own opinion, are physicians able to assess functional limitations based on a written DASI report?
5. What are the characteristics of the DASI in daily practice?
6. What is the patient satisfaction evaluation for physicians who conduct a DASI interview?
7. What comments on the DASI do insurance physicians have?
8. What is the opinion about using the DASI that insurance have?

9. What is the intra- and inter-rater reliability of functional limitations assessments using the DASI?
10. What is the content and concurrent validity of functional limitation assessments using the DASI?

Outline of this thesis

In **Chapter 2**, a model is presented in which the process of disability assessment, the instruments used and the role of the assessor is addressed. On the basis of this model, the causes of inter-rater variability and suggestions for improvement are discussed.

In **Chapter 3**, a systematic review of the literature is presented, which comprises studies on instruments for assessing functional limitations in workers' compensation claimants, and their psychometric properties.

In **Chapter 4**, a study is described which investigated whether the provision of detailed information concerning participation and activity limitations, as compared to medical information alone, influences the assessment of work limitations by physicians. Three different groups of insurance physicians were given different kinds of information on the same 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 forms of information. Agreement percentages within the groups and differences between the groups as to scores given on the work limitation items of the Functional Ability List (FAL) were measured.

In **Chapter 5**, a study is described in which written reports of DASI interviews are used to investigate whether physicians are able to perform a disability assessment based on a written report. In addition, the inter-rater reliability among physicians was measured by computing the percentage agreement with respect to the mental and physical items of the Functional Information System (FIS) and the Mental Ability List (MAL).

In **Chapter 6**, a study is described in which video recordings of DASI interviews were used to analyze the content of DASI interviews; physicians were asked for their comments about the interviews and to determine inter- and intra-rater reliability of assessments. The interviews were analyzed by measuring the duration of the different topics of the interviews. The inter-rater reliability among the physicians was measured by computing the percentage agreement with respect to the mental and physical items of the Functional Information System (FIS) and the Mental Ability List (MAL). To measure intra-rater reliability, the insurance physicians who made the recordings were asked to fill in the FIS and MAL right after the recordings and after seeing the video again six months later.

In **Chapter 7**, a randomized controlled trial is described in which employees applying for a work-disability pension were independently interviewed and examined either by two physicians who had completed DASI training or by two physicians from a control group without any training. Agreement percentages within both groups of physicians, eligibility for a disability benefit, and differences between the groups in terms of the scores given on the work-limitation items from FAL were measured to investigate reliability and concurrent validity. To determine the content validity, the insurance physicians who completed DASI training were asked to fill out a questionnaire concerning their opinion of the DASI. Additionally, patients filled out a questionnaire to measure their satisfaction as to the behavioral aspects of the physicians.

In **Chapter 8**, the main findings of the studies are presented and discussed.

In **the Addendum** an extensive description of the FAL and the DASI can be found.

Reference List

1. Functional Ability List (FAL). Claim Beoordelings- en BorgingsSysteem (CBBS). [Claim Assessment and Monitoring System]. Amsterdam: LISV, 2002.
2. Matheson LN, Kaskutas V, McCowan S, Shaw H, Webb C. Development of a database of functional assessment measures related to work disability. J Occup Rehabil 2001; 11(3):177-199.
3. Innes E. Reliability and Validity of Functional Capacity Evaluations: An Update. International Journal of Disability Management Research 2006; 1(1):135-148.
4. 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(1):130-140.
5. Williams RM, Schmuck G, Allwood S, Sanchez M, Shea R, Wark G. Psychometric evaluation of health-related work outcome measures for musculoskeletal disorders: a systematic review. J Occup Rehabil 2007; 17(3):504-521.
6. Söderberg E, Alexanderson K. Sickness certification practices of physicians: a review of the literature. Scandinavian Journal of Public Health 2003; 31(6):460-474.
7. Mayer J, Mooney V, Matheson L, Leggett S, Verna J, Balourdas G et al. Reliability and validity of a new computer-administered pictorial activity and task sort. J Occup Rehabil 2005; 15(2):203-213.
8. Spanjer J. De betrouwbaarheid van WAO beoordelingen, onderzoek. [The inter- and intra- rater reliability of disability pension assessments]. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2001; 9:234-241.
9. Brouwer S, Dijkstra PU, Gerrits EHJ, Schellekens JMH, Groothoff JW, Geertzen JHB et al. Intra- en interbeoordelaarsbetrouwbaarheid 'FIS-belastbaarheidspatroon'en 'Functionele Mogelijkheden Lijst'. [Intra- and inter-rater reliability 'Functional Information System' and 'Functional Ability List']. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2002; 11:360-367.
10. Ehrlich GE, Wolfe F. On the difficulties of disability and its determination. Rheumatic Diseases Clinics of North America 1996; 22(3):613-621.
11. Loeser JD, Sullivan M. Doctors, diagnosis, and disability: a disastrous diversion. Clinical Orthopaedics & Related Research 1997; 336:61-66.
12. Hunt DG, Zuberbier OA, Kozlowski AJ, Berkowitz J, Schultz IZ, Milner RA et al. Are components of a comprehensive medical assessment predictive of work disability after an episode of occupational low back trouble? Spine 2002; 27(23):2715-2719.
13. Rainville J, Pransky G, Indahl A, Mayer EK. The physician as disability advisor for patients with musculo-skeletal complaints. Spine 2005; 30(22):2579-2584.
14. Tait RC, Chibnall JT, Andresen EM, Hadler NM. Disability determination: validity with occupational low back pain. Journal of Pain 2006; 7(12):951-957.
15. Mul CAM, Verkleij HEM, Hazeltet AM, Roos F de. Kwaliteit van de oordeelvorming. [Quality of assessment]. Hoofddorp: TNO, 1999.
16. 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 assessment]. Zoetermeer: CTSV, 2000.
17. Boer WEL de, Wijers JHL, Spanjer J, van der Beijl I, Zuidam W, Venema A. Gespreksmodellen in de verzekeringsgeneeskunde. [Interview models in insurance medicine]. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2006; 14:17-23.
18. Boer WEL de, Duin JA, Herngreen H. Handleiding Methodisch beoordelingsgesprek. [Manual Methodical Assessment Interview]. Utrecht: NSPOH: 1997.
19. Boer WEL de, Meijers JM, Verhoef M. De Multi Causale Analyse: theorie en praktijk. [Multi-causal Analysis, theory and practice]. Hoofddorp: TNO Arbeid, 2000.

20. Spanjer J. Belastbaarheidsgerichte beoordeling, een onderzoek naar intra- en interbeoordelaars betrouwbaarheid van WAO beoordelingen. [Disability Assessment Structured Interview, a study of inter- and intra-rater reliability of disability pension assessments]. Amsterdam: LISV, 2000.
21. International classification of functioning, disability and health. Geneva: World Health Organisation, 2001.
22. Innes E, Straker L. Reliability of work-related assessments. Work 1999; 13(2):107-124.
23. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. 2 ed. Upper Saddle River, New Jersey, Prentice Hall, Inc, 2000.
24. Innes E, Straker L. Validity of work-related assessments. Work 1999; 13(2):125-152.

CHAPTER 2

Sources of variation in work disability assessment

Jerry Spanjer
Boudien Krol
Sandra Brouwer
Johan W. Groothoff

Work. In press

ABSTRACT

The assessment of work disability due to health problems is a difficult task because there is no straightforward relationship between disease and disability. As a result, there is wide inter-rater variability between physicians in assessing work disability. The aim of this paper is to discuss the sources of the inter-rater variability and to describe possibilities for its reduction. A model is presented in which the process of disability assessment, the instruments used and the role of the assessor is addressed. On the basis of this model, the causes of inter-rater variability and suggestions for improvement are discussed.

INTRODUCTION

Assessment of work disability – defined in the present paper as limitations in performing work activities due to a given health problem – is an important task because of the serious personal and social consequences of the assessor's decision. In most countries work disability assessment is the task of a physician[1]. However, research indicates that physicians often overestimate disability[2;3] and there can be wide inter-rater variability between physicians when assessing work disability[4-7]. There are several explanations for the variability between physicians, for instance:

- the level of functioning of a patient with a particular disease cannot be measured by physical examination or deduced from the diagnosis itself, often disease and disability are only loosely connected[8-12];
- in work disability, pain and fatigue are frequently the only symptoms, thus lacking further specific pathology. For instance, in most low back pain patients there are no clinical findings and specific diagnoses can only be made for fewer than 10%[13]. The issue is to determine under which circumstances pain – which is associated with a certain activity – makes performing that activity impossible. Without clear organ impairment there is no medical basis to consider a patient's behaviour as a disability. Therefore, disability appears to be a social and a medical concept.

Since work disability cannot simply be measured or readily determined from a diagnosis, special attention to the assessment process is required. The first step in the process of assessing is the collection of information, which can be performed in several ways. Besides clinical evaluation, there are other methods to assess work disability, such as self-evaluation by means of questionnaires, interview procedures and performance tests. Outcomes based on these methods vary considerably – self-reported disability is often more severe than disability assessed by a physician, whereas performance tests indicate the least disability[14-19].

Subsequently, the information collected from diagnoses, physical examinations, laboratory findings, workload, patient behaviour and the patients' own opinions have to be weighed and documented.

The framework in which disability assessment takes place may also be an important factor. Disabled patients applying for a job may emphasize their capabilities in order to get the job. Patients in rehabilitation may want to show their progress in therapy. However, when claiming disability benefits patients may tend to emphasize their disabilities in order to qualify for a work disability pension.

Due to all these aspects, assessment of work disability is complex, which causes variation in outcomes between assessors. Research into assessment methods is difficult to accomplish because of the lack of a proper benchmark or "gold standard". Therefore, the aim of this paper is to discuss the sources which may contribute to inter-rater variability in the assessment of work disability, and to describe how the variability can be reduced by providing suggestions for improvement.

To describe the different sources of variability we constructed a model where the process of disability assessment and the instruments used are detailed (Figure 1). Three steps can be distinguished – collection of information, interpretation and documentation. Each of these steps, including the instruments used, will be described in detail and the role of the assessor will be discussed.

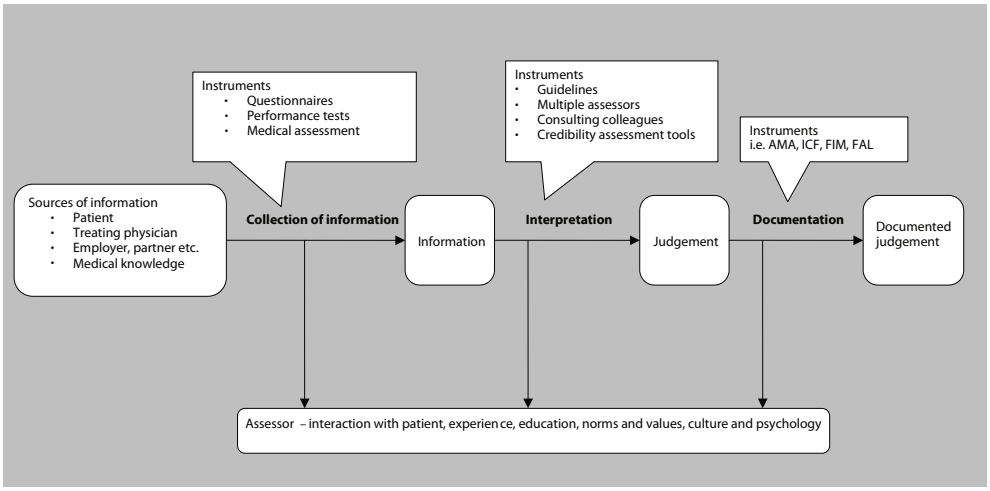

Figure I. A Model for Disability Assessment

AMA = American Medical Association, ICF = International Classification of Functioning, FIM = Functional Independence Measure, FAL = Functional Ability List

COLLECTION OF INFORMATION

Information can be collected from several sources. These include the patient, the treating physician and the patient's employer or partner. Information obtained from family or a partner may be important when the patient is unable to provide the information him or herself. It is important to know the source of the information to estimate its merit.

Because assessment concerns patients with medical conditions, clinical information is needed first. This can be acquired by taking a standard medical history, physical examination, imaging and laboratory studies. In general however, this information is often already available and in the hands of the treating physician. Medical information is helpful in knowing the cause of the disability, while information on diagnosis, treatment and medication can shed light on the disability itself. However, treating physicians are often hesitant to provide information on disability because it can frustrate their relationship with their patients and because they lack the skill to evaluate disability as it impacts on work capacity[20-22]. Therefore, the information on disability and its social implications needs to be collected by other means.

Questionnaires, performance tests and medical assessment are all instruments that can be used to collect information on disability.

Questionnaires

The use of questionnaires allows the structuring of questions on patient disabilities. Questionnaires can be filled out by the patients themselves or by experts during a structured interview. Self-report questionnaires often concern a specific disease such as migraine[23] or low back pain[24], but can also be administered for more general use, for example, the 'Work Limitations Questionnaire'[25] or the 'Pain Disability Questionnaire'[26]. The validity and reliability of self-report questionnaires can be limited. For instance, a systematic review of questionnaires designed for assessing functional status in low back pain patients showed that in 36 questionnaires only a few could be considered acceptably validated and that considerable variation with respect to their main concept and content was found[27].

Expert questionnaires follow a standard format for structured interviews, with a trained interviewer asking questions to prevent patient bias. An example of this data collection procedure is the 'Functional Status Examination'[28]. Generally, questionnaires only register what the patient reports, without an actual assessment.

Performance tests

Performance tests or functional capacity evaluation (FCE) measure the actual performance of the patient in a research centre. Several tests can be performed, such as the lifting of different weights or working above shoulder height. FCE's are commonly used to measure physical abilities in work-related tasks in patients with musculoskeletal disorders. Numerous studies on the psychometric qualities of FCE's have been conducted. Systematic reviews conclude that, although the reliability often is satisfactory, the validity of the results is questionable because the sincerity of effort, ability to perform work outside a laboratory setting and the prediction of injury are difficult to measure[29-31].

Medical assessment

In many countries a medical assessor has a face-to-face interview with the patient when assessing disability for a disability benefit decision. An interview with the patient is often used in the collection of information concerning work disability[1], but to the best of our knowledge, no international literature on the structure of these interviews is available. The interview can be fully structured, as in expert questionnaires, or a looser structure can be used. Standard medical history taking and physical examination alone are not sufficient to assess work disability[10;22;32]. Specific information on activity limitations and participation is useful compared to medical history-taking alone[33;34]. In the Netherlands, interview protocols play an important role in the medical assessment of disability for social security benefit decisions[1]. Three Dutch semi-structured interview protocols are described[35]. In addition to a medical history and the registration of complaints, the main topics in these interviews are functioning in daily life and work, the opinion of the patient about disabilities and possibilities,

the description of a usual day and detailed examples of the disabilities experienced. Conversation and interrogation skills, and the observation of consistency in the patient's story are important tools. Observation and physical medical examination can contribute to disability assessment. Research indicates that it is possible to obtain high inter-observer agreement when observing the patient[36]. However, observed disability levels are lower than self-reported disability[18] or even lower than that predicted by clinicians[2]. Physical medical examination may have low inter-observer agreement and is not necessarily related to disability[37]. For instance, observation may reveal that a patient is able to lift heavy weights despite there being considerable functional limitations revealed by the physical examination of their back.

Other information collection instruments are, for instance, literature searches for information on disability associated with a specific disease, and expert consultation.

INTERPRETATION

After information has been collected a decision has to be made as to the patient's functional (dis)abilities. Because the disability the patient claims and the patient's behaviour do not always correspond to objective medical findings, this is a difficult task. For example, is a patient disabled if all the patient presents is non-specific low back pain or chronic fatigue syndrome unsupported by any objective medical findings? In other words, when is the pain associated with performing certain tasks severe enough to be regarded as incapacitating? Is a patient, testing normal for exercise after a myocardial infarction, medically restricted if he or she rests every afternoon because of fatigue?

Several instruments can help to answer these questions and obtain agreement between assessors. Using guidelines and protocols, physicians can reach mutual agreement on these questions to prevent differences in judgement. Physicians and other professionals can determine how the information collected should be interpreted. For instance, in the Netherlands the 'medical disability criterion' is legally documented[38;39]. According to these criteria, the patient's experience should be the starting point of the assessment. However, a loss of autonomy must be present, along with the requirement that impairments, activity limitations and handicaps should fit together consistently.

Regrettably, these guidelines and laws cannot prevent differences between assessors. In fact, laws and guidelines can be interpreted differently and professionals do not always follow prescribed guidelines[40]. Furthermore, the contents of guidelines are not always sufficiently well known. In addition, given the complexity of what is sought to be measured, not every patient or situation will fit within existing guidelines. Moreover, disability assessment often is less a technical matter than a normative one but guidelines are based on formal rationality and deny the normative dimension[41].

Another way of reducing variation in disability judgements is multiple assessors duplicating the same disability assessment. The same patient can be assessed by two or more physicians independently or one physician in consultation with a colleague. Educational meetings about themes such as 'Disability assessment in diseases without objective medical findings' may decrease

variation between physicians.

Credibility assessment tools are meant to test whether the information provided is reliable and true. The polygraph is a well-known example, though its use in disability assessment is unacceptable in most cultures. Another more promising tool is the 'Statement validity assessment' for measuring the credibility of verbal statements[42]. Content criteria are described for statements to evaluate their reliability – for example, logical structure, the quantity of detail, unusual details and contextual embedding.

DOCUMENTATION

Once the information has been collected and interpreted, the assessor has to document the assessment. This can be a written report of the presumed disability level, though instruments for recording disability items are also used. However, the majority of these latter instruments do not document work disability, for instance:

- in rehabilitation medicine ADL items such as 'toileting' and 'dressing' can be more important than disability for work, for example, 'The Functional Independence Measure'[43];
- instruments that measure impairments rather than disability, such as the AMA guide[44];
- instruments that register disability claimed by a patient rather than an expert's assessment, such as the 'WHO Disability Assessment Schedule'[45];
- the 'International Classification of Functioning, Disability and Health' not only registers work and ADL disability, but also impairments and the personal factors involved[46].

Instruments that record work disability contain items on the demands of the workplace. As an example, two items of the Dutch 'Functional Ability List' are presented in Table 1[47].

Different countries use different instruments. Studies on their validity and reliability are often unavailable. Problems such as poor operational definitions, limited domains of functioning, low ceiling effects and the combination of impairments and work disability are frequently present.

Table I. Two examples of 'Functional Ability List' items

Item	Description
Lifting or carrying	0 normal, can carry or lift about 15 kg (toddler) 1 can carry or lift about 10 kg (small toddler) 2 limited, can carry or lift about 5kg (bag of potatoes) 3 severely limited, can carry or lift about 1kg (one litre of milk)
Sitting	0 normal, can sit for about 2 hours (car trip) 1 slightly limited, can sit for about one hour on end (movie) 2 limited, can sit for about half an one hour on end (meal) 3 severely limited, can sit for less than 15 minutes on end (news bulletin)

THE ASSESSOR

An important factor in assessing work disability is the assessor, playing a major role in all three steps of assessment – collection of information, interpretation and documentation. During the assessment procedure interaction with the patient is of utmost importance. In the case of language problems, mutual communication can be difficult to achieve and the collection of sufficient information may become a problem. Research indicates that where communication problems are encountered the assessor tends to stereotype[48]. Experience and education is important, an assessor with better conversation skills can collect more information and more relevant information. Moreover, the cultural background, norms and values of the assessor all play a role. For example, research has shown that independent medical examiners assess lower levels of disability than treating physicians due to differences in opinion rather than skills or training[3].

In assessing work disability several psychological mechanisms play a role. Examples include the 'rank-effect' – that is, previous assessment influences the subsequent assessment – and 'confirmation bias' – the tendency to search for or interpret new information in such a way as to confirm preconceptions and overlook information and interpretations conflicting with prior beliefs[49].

DISCUSSION

Assessment of work disability is a delicate and important task since the financial and social consequences for the patient and society as a whole are considerable. Therefore, differences in outcome in terms of the level of work disability between assessors is unwanted and should be reduced.

The difference in outcome is mainly caused by the fact that assessment of work disability is not an exact science but rather a subjective judgement. Subjective information from the patient is often an important and indispensable component in forming such judgements, but can be coloured because the patient has a certain interest in the result of the assessment. Moreover, the assessing physician is subjective in collecting and interpreting information. A key question is to what extent a chronically ill patient's behaviour needs to be assessed as a disability. Perhaps a lesson concerning the assessment process can be learned from the law, where rules are written down and jurisprudence and precedents are used to narrow the scope for interpretation. Judgement in aesthetic sports, such as figure skating, where multiple judges and computerized assistance are used to assess the performance of skaters, can also serve as an example.

Considering the possible sources of variation between assessors, which are described in Figure 1, a number of suggestions for the assessment of work disability with reduced inter-rater variance can be offered:

- the use of a fixed, structured information gathering method so as to base the judgement of different assessors on the same information;

- the use of clear guidelines to interpret the information, and monitoring to ensure implementation of these guidelines;
- the use of multiple assessors trained in the assessment of work disability;
- the use of validated and reliable instruments to document the assessed work disability.

CONCLUSION

Work disability assessment is a delicate and complex task. The consequences of assessment for the patient and for society as a whole may be far-reaching. Therefore, the wide inter-rater variability between physicians that seems to exist when assessing work disability is undesirable. In order to reduce variability in outcome we propose making the assessment process transparent by standardizing information collection, using clear guidelines and a reliable and validated instrument to document the assessment, and using multiple and specially trained assessors.

Reference List

1 De Boer WEL, Besseling JJM, Willems JHBM. Organisation of disability evaluation in 15 countries. Pratiques et Organisations de Soins 2007;38:205-217.

2 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. Archives of Physical Medicine & Rehabilitation 2004;85:837-839.

3 Lax MB, Manetti FA, Klein RA. Medical evaluation of work-related illness: evaluations by a treating occupational medicine specialist and by independent medical examiners compared. International Journal of Occupational & Environmental Health 2004;10:1-12.

4 Innes E, Straker L. Reliability of work-related assessments. Work 1999;13:107-124.

5 Wolfson AM, Doctor JN, Burns SP. Clinician judgments of functional outcomes: how bias and perceived accuracy affect rating. Archives of Physical Medicine & Rehabilitation 2000;81:1567-1574.

6 Spanjer J. De reproduceerbaarheid van WAO-beoordelingen, een literatuuronderzoek. [Reproducibility in disability assessment: a literature search]. Tijdschrift voor bedrijfs- en verzekeringsgeneeskunde 2001;9:195-8.

7 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;11:360-7.

8 Ehrlich GE, Wolfe F. On the difficulties of disability and its determination. Rheumatic Diseases Clinics Of North America 1996;22: 613-621.

9 Loeser JD, Sullivan M. Doctors, diagnosis, and disability: a disastrous diversion. Clinical Orthopaedics & Related Research 1997;336:61-66.

10 Hunt DG, Zuberbier OA, Kozlowski AJ, Berkowitz J, Schultz IZ, Milner RA , et al. Are components of a comprehensive medical assessment predictive of work disability after an episode of occupational low back trouble? Spine 2002;27:2715-2719.

11 Rainville J, Pransky G, Indahl A, Mayer EK. The physician as disability advisor for patients with musculoskeletal complaints. Spine 2005;30:2579-2584.

12 Tait RC, Chibnall JT, Andresen EM, Hadler,N.M. Disability determination: validity with occupational low back pain, Journal of Pain 2006;7:951-957.

13 Deyo RA, Rainville J, Kent DL. What can the history and physical examination tell us about low back pain? JAMA 1992;268: 760-765.

14 Wijlhuizen GJ, Ooijendijk W. Measuring disability, the agreement between self evaluation and observation of performance. Disability & Rehabilitation 1999;21:61-67.

15 Lee CE, Simmonds MJ, Novy DM, Jones S. Self-reports and clinician-measured physical function among patients with low back pain: a comparison. Archives of Physical Medicine & Rehabilitation 2001;82: 227-231.

16 Rogers JC, Holm MB, Beach S, Schulz R, Cipriani J, Fox A, Starz TW. Concordance of four methods of disability assessment using performance in the home as the criterion method. Arthritis & Rheumatism: Arthritis Care & Research 2003;49:640-647.

17 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. Disability & Rehabilitation 2005;27: 999-1005.

18 Dowrick AS, Gabbe BJ, Williamson OD, Wolfe R, Cameron PA. A comparison of self-reported and independently observed disability in an orthopedic trauma population. Journal of Trauma 2006;61:1447-1452.

19 Hart DL, Kirk M, Howar J, Mongeon S. Association between clinician-assessed lifting ability and workplace tolerance and patient self-reported pain and disability following work conditioning, Work 2007;28:111-119.

20 Pransky G, Katz JN, Benjamin K, Himmelstein J. Improving the physician role in evaluating work ability and managing disability: a survey of primary care practitioners. Disability And Rehabilitation 2002;24:867-874.

21 Söderberg E, Alexanderson K. Sickness certification practices of physicians: a review of the literature. Scandinavian Journal of Public Health 2003;31:460-474.

22 Wahlstrom R, Alexanderson K. Chapter 11. Physicians' sick-listing practices, Scandinavian Journal of Public Health 2004;32:222-255.

23 Stewart WF, Lipton RB, Whyte J, Dowson A, Kolodner K, Liberman JN, Sawyer J. An international study to assess reliability of the Migraine Disability Assessment (MIDAS) score. Neurology 1999;53:988-994.

24 Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. Physical Therapy 2002;82: 8-24.

25 Lerner D, Amick BC III, Rogers WH, Malspeis S, Bungay K, Cynn D. The Work Limitations Questionnaire, Medical Care 2001;39:72-85.

26 Anagnostis C, Gatchel RJ, and Mayer TG. The Pain Disability Questionnaire: a new psychometrically sound measure for chronic musculoskeletal disorders... including commentary by Fairbank JCT. Spine 2004;29:2290-2303.

27 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.

28 Dikmen S, Machamer J, Miller B, Doctor J, Temkin N. Functional status examination: a new instrument for assessing outcome in traumatic brain injury. Journal Of Neurotrauma 2001;18:127-140.

29 Gouttebarge V, Wind H, Kuijer PP, Frings-Dresen MHW. Reliability and validity of Functional Capacity Evaluation methods: a systematic review with reference to Blankenship system, Ergos work simulator, Ergo-Kit and Isernhagen work system. International Archives Of Occupational And Environmental Health 2004;77:527-537.

30 Pransky GS, Dempsey PG. Practical aspects of functional capacity evaluations. J Occup Rehabil. 2004;14:217-229.

31 Innes E. Reliability and validity of functional capacity evaluations: an update. International Journal of Disability management Research 2006;1:135-148.

32 Swartling MS, Hagberg J, Alexanderson K, Wahlstrom RA. Sick-listing as a psychosocial work problem: a survey of 3997 Swedish physicians. J.Occup.Rehabil. 2007;17:398-408.

33 Spanjer J, Krol B, Brouwer S, Groothoff JW. Inter-rater reliability in disability assessment based on a semi-structured interview report. Disability and Rehabilitation 2008;30:1885-1890.

34 Spanjer J, Krol B, Popping R, Groothoff JW, Brouwer S. Disability assessment interview: the role of detailed information on functioning in addition to medical history-taking. J Rehabil Med 2009;41:267-272.

35 De Boer WEL, Wijers JHL, Spanjer J, van der Beijl I., Zuidam W, Venema A. [Interview models in insurance medicine], Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2006;14:17-23.

36 Koho P, Aho S, Watson P, Hurri H. Assessment of chronic pain behaviour: reliability of the method and its relationship with perceived disability, physical impairment and function. Journal of Rehabilitation

Medicine 2001;33:128-132.

37 Horneij E, Hemborg B, Johnsson B, Ekdahl C. Clinical tests on impairment level related to low back pain: a study of test reliability. Journal Of Rehabilitation Medicine 2002;34:176-182.

38 Berg M, Horstman K, Plass S, van Heusden M. Guidelines, professionals and the production of objectivity: standardisation and the professionalism of insurance medicine. Sociology Of Health & Illness 2000;22:765-791.

39 Schattingsbesluit arbeidsongeschiktheidswetten. [Assessment Decree Occupational Disability Act]. The Hague, Netherlands: Ministry of Social Affairs and Employment, 2004.

40 Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PAC, Rubin HR. Why Don't Physicians Follow Clinical Practice Guidelines? A Framework for Improvement. JAMA 1999;282:1458-1465.

41 Meershoek A, Krumeich A, Vos R. Judging without criteria? Sickness certification in Dutch disability schemes. Sociology Of Health & Illness 2007;29:497-514.

42 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 and Criminological Psychology 2005;10:225-245.

43 Keith RA, Granger CV, Hamilton BB, Sherwin FS. The functional independence measure: a new tool for rehabilitation. Adv Clin Rehabil 1987;1:6-18.

44 Gloss DS, Wardle MG. Reliability and validity of American Medical Association's guide to ratings of permanent impairment. JAMA 1982;248: 2292-2296.

45 World Health Organisation Disability Assessment Schedule (WHODAS II).Geneva: WHO, 2000.

46 International classification of functioning, disability and health. Geneva: WHO, 2001.

47 Functionele Mogelijkheden Lijst (FML). [Functional Ability List (FAL)]. Claim Beoordelings- en BorgingsSysteem (CBBS). Amsterdam: LISV, 2002.

48 Meershoek A, Krumeich A, Desain L. Arbeidsongeschiktheid, reintegratie & etniciteit. [Work disability, reintegration and ethnicity]. Apeldoorn: Het Spinhuis, 2005.

49 Bornstein BH, Emler AC. Rationality in medical decision making: a review of the literature on doctors' decision-making biases. Journal Of Evaluation In Clinical Practice 2001;7:97-107.

CHAPTER 3

Instruments used to assess functional limitations in workers' compensation claimants: a systematic review

Jerry Spanjer
Sandra Brouwer
Johan W. Groothoff

Submitted for publication

ABSTRACT

Objectives

First, to systematically review the literature on instruments for assessing functional limitations in workers' compensation claimants. Second, to review the psychometric properties for those instruments found.

Methods

Electronic searches of Medline, Embase, CINAHL and PsycINFO for literature between 1980 and December 2008 were performed to identify studies focusing on the psychometric properties of instruments used to assess functional limitations in workers' compensation claimants. Two independent reviewers applied the inclusion criteria to select relevant articles and then evaluated the psychometric qualities of the instruments found.

Results

Of the 712 articles that were identified, ten studies met the inclusion criteria, reporting on four different instruments: the Roland-Morris Disability Questionnaire (RDQ), the Patient-Specific Functional Scale (PSFS), the Isernhagen Work System (IWS) and the Multiperspective Multidimensional Pain Assessment Protocol (MMPAP). For all four of these instruments, predictive validity was assessed, for three of them construct validity, and for none of them content validity or reliability. The questionnaires (RDQ and PSFS) did not focus specifically on the work situation and measured functional limitations in a limited manner. The psychometric qualities of the IWS were poor to moderate. For the MMPAP, only predictive validity was measured. The instruments assessed three to 34 physical functional limitations, and no instruments that assessed mental limitations in claimants were found.

Conclusion

We did not find any instruments with satisfactory psychometric qualities for assessing functional limitations in claimants. Further development and research into evidence-based instruments for assessing functional limitations in claimants for workers' compensation is needed, given the importance of these assessments.

INTRODUCTION

The assessment of functional limitations is an important part of the evaluation for the work disability pension and has immense individual, financial and social consequences. When workers become disabled in the Netherlands, their employers still have to pay wages for two years. Then a workers' compensation benefit from the Social Security Office can be applied for. The benefit is based on the loss of the wage-earning capacity of the claimant. An insurance physician first assesses the functional limitations, then a labor expert assesses what the claimant theoretically is still able to earn in suitable work. Work disability is a multifactorial phenomenon, influenced by personal (physical and mental status) and environmental factors[1-3]. According to the International Classification of Functioning, Disability and Health (ICF), limitations in functioning are defined as the limitations in performing a task or action by an individual. It is an umbrella term encompassing all body functions, activities and participation[4]. In the present study, the definition of functional limitations is: limitations in or inability to perform certain physical activities such as walking and lifting, or mental activities such as concentrating and conflict handling. Therefore, functional limitations can be distinguished from symptoms (such as pain and fatigue), activity limitations (such as self-care tasks and gardening) and participation restrictions (such as leisure time activities and work).

Instruments used to assess functional limitations have to provide reliable and valid information to enable appropriate decisions about the work disability pension[5;6]. Reliability is the extent to which an instrument is free from error and consistent over time, between different raters or between parts of the test[6]. Validity is considered to be the extent to which an instrument measures what it is intended to measure[5]. One validates not an instrument *per se*, but rather it's use in a specific setting or in a specific target group[7;8]. For instance, an instrument can be valid in a rehabilitation setting but not in the assessment of a workers' compensation benefit. If workers believe a certain test could affect their ability to receive benefits, this may affect outcomes[9;10]. Instruments that measure physical limitations may not be appropriate for measuring mental limitations.

In the Netherlands, the insurance physician bases the assessment of the functional limitations on an interview with the claimant, a medical examination, and information from treating physicians; and uses no other specific instruments. The reliability and validity of these assessments are questionable[11].

Instruments for assessing functional limitations which are described in the literature[12;13] include functional capacity evaluations[14], questionnaires[15;16] and expert assessments by physicians[17]. Functional capacity evaluations (FCE) are standardized tests which measure actual physical performance, and are used to determine a subject's ability to perform work-related activities[18]. Several reviews can be found on the psychometric properties of these instruments[12;16;19;20], but there are some concerns as to their usefulness in assessing functional limitations in workers' compensation claimants:

- several instruments have been studied in a rehabilitation setting or in job fitness for healthy people,

which makes their validity in claimants for workers' compensation questionable.
- questionnaires often do not have a work-related point of reference: they consider activities, such as self-care tasks, domestic tasks and sickness absence; but whether the patient can lift 10 kg at work remains unknown.
- the items of the instruments often inquire about a combination of symptoms, functional limitations, and about activity and participation restrictions, rather than focusing only on functional limitations[21].
Until now, no reviews have been published concerning instruments that assess mental and physical functional limitations in claimants for workers' compensation. To fill this gap, we systematically reviewed the literature on instruments for assessing functional limitations in claimants for workers' compensation. For those instruments found, we evaluated the psychometric properties, such as reliability and validity.

METHODS

Literature search
Studies were identified by searches of the electronic bibliographic databases Medline (biomedical literature), Embase (biomedical and pharmacological literature), CINAHL (nursing and allied health literature) and PsycINFO (psychological literature). The searches were limited to literature published between 1980 and December 2008. The search terms used were: *Disability evaluation [Major topic]; AND Observer variation [Mesh] OR Psychometrics [Mesh] OR Reproducibility of Results [Mesh] OR reliability OR validity; AND Workers' Compensation [Mesh] OR Work* OR claimants OR job* OR occupation* OR vocation**.

Articles were included in the review if the following criteria were met:

1. An instrument was described for assessing functional limitations in a work setting.
2. The instrument was used on claimants for a workers' compensation benefit.
3. The article was published in English, German or Dutch.
4. The article was a primary peer-reviewed research study.
5. Psychometric properties of the instrument were presented.

Excluded were studies regarding return to work without assessing functional limitations, job fitness in healthy people, malingering and clinical setting. We also excluded case studies, letters to the editor and book chapters. Review papers were only used to screen for further original papers. References of retrieved articles were screened for additional relevant studies.
Applying these criteria, the first two authors (JS, SB) independently reviewed the titles and abstracts of the literature to identify potentially relevant articles. If the title and abstract did not provide enough information to decide whether or not the inclusion criteria were met, the article was included for full-text selection. From the articles included, we read the full text and the same two

reviewers applied the inclusion criteria to the full text. Disagreements between the reviewers were discussed and resolved during a consensus meeting.

The names of identified instruments were used as terms for a further search of the electronic databases. We systematically reviewed the literature on their reliability and validity. To identify eligible studies we used the above-mentioned electronic bibliographic databases with the following keywords: *"the name of the instrument"; AND Psychometrics [Mesh] OR Reproducibility of Results [Mesh] OR reliability OR validity*.

Table 1. Quality criteria used for psychometric properties[6;22]

Property	Definition	Quality criteria
1. Content validity	The extent to which the domain of interest is comprehensively sampled by the items in the test	poor: the test does not measure what it is intended to measure moderate: not all relevant components are included good: the test measures all relevant components
2. Internal consistency	The extent to which items in a (sub)scale are intercorrelated, thus measuring the same construct	Cronbach's α ≤ 0.70 poor; 0.71-0.80 moderate; >0.80 good
3. Criterion validity	The extent to which scores on a particular test relate to another valued measure	Correlation between test and the criterion measure r ≤ 0.50 poor; 0.51 – 0.75 moderate; > 0.75 good
4. Construct validity	The extent to which a test is well correlated with a hypothetical construct or theoretical expectation	> 75% of the results are in accordance with the theoretical expectation. Convergence between tests: poor r ≤ 0.30; moderate 0.31-0.60; good >0.60
5. Reproducibility		
5.1. Agreement	The extent to which the scores on repeated measures are close to each other (absolute measurement error)	MIC < SDC OR MIC outside the LOA OR convincing arguments that agreement is acceptable
5.2. Reliability	The extent to which patients can be distinguished from each other, despite measurement errors (relative measurement error)	ICC: poor < 0.75; moderate 0.75 – 0.90; good > 0.90 Kappa: poor < 0.40; moderate 0.40 – 0.60; good > 0.60
6. Responsiveness	The ability of a test to detect clinically important changes over time	Significant difference in t-test: p< 0.05 Effect Size and Standard Response Mean: poor: 0.20 – 0.50; moderate 0.51-0.80; large >0.80
7. Floor and ceiling effects	The number of respondents who achieved the lowest or highest possible score	≤ 15% of the respondents achieved the highest or lowest possible scores

MIC = minimal important change; SDC = smallest detectable change; LOA = limits of agreement; ICC = intraclass correlation.

Quality assessment

The psychometric properties of the instruments were assessed by the first two authors independently. Analysis and interpretation of content validity, internal consistency, criterion validity, construct validity, reproducibility (agreement and reliability), responsiveness, and floor and ceiling effects were rated using the criteria described by Terwee et al.[22] and Innes and Straker[5;6] (Table 1). Possible ratings for the psychometric properties were good (+), moderate (±) and poor (-).

Description of the instruments

Descriptive data for the instruments found were extracted from the publications, and included the type of instrument (questionnaire, performance test, physician assessment), time to complete, number of items, number of scales, and the target population for which the instrument had been developed.

RESULTS

The search in Medline yielded 439 articles, CINAHL 174, PsycINFO 17 and Embase 281. From this total of 911 articles, 712 remained after removal of duplicates. Reference tracing resulted in one additional article. A total of 676 articles were read by title and abstract only, 37 articles were selected for full text reading. The most important reason for excluding studies was because they were not conducted in a claim setting but in a rehabilitation setting or did not measure functional limitations. Fifteen articles were not published in English, German or Dutch. A total of ten studies fulfilled the inclusion criteria, reporting on four different instruments: the Roland-Morris Disability Questionnaire (RDQ)[23;24], the Patient-Specific Functional Scale (PSFS)[25], the Isernhagen Work System (IWS)[26-31], and the Multiperspective Multidimensional Pain Assessment Protocol (MMPAP)[32].

An overview of these studies and the psychometric properties of the instruments are described in Table 2. For all instruments, the predictive validity was described; for all instruments but the MMPAP, construct validity was studied. The RDQ was the only instrument for which responsiveness, internal consistency and floor and ceiling effects were measured. None of the studies examined the face and content validity or reproducibility of the instruments in workers' compensation claims.

Below we will describe the content of the four instruments and their psychometric properties investigated in a claim setting (Table 3).

Roland-Morris Disability Questionnaire (RDQ)

The RDQ[33] is a questionnaire that is widely used to assess self-reported physical disability associated with low back pain. The 24 dichotomous items cover symptoms (4 items, e.g., pain and poor appetite), activity limitations and participation restrictions (12 items, e.g., staying in bed and avoiding heavy work), and functional limitations (8 items, e.g., walking, standing and climbing stairs). Total scores

Table 2. Studies describing instruments for assessing functional limitations in claimants for workers' compensation

Instrument	Population	Psychometrics	Outcome	1st author, year
Roland-Morris Disability Questionnaire (RDQ)	N = 284 D = back pain T = median 56 days	Internal consistency Floor and ceiling effects Responsiveness Construct validity (SF-12, SF-36)	+ + ±/+ +	Turner[24], 2003
	N = 959 D = back pain T = 0 – 90 days	Predictive validity (return to work) Construct validity (SF-12, NRS-101)	+ +	Baldwin[25], 2007
Patient-Specific Functional Scale (PSFS)	N = 294 D = musculoskeletal disorder T = mean 388 days	Predictive validity (time receiving benefit; time to claim closure; recurrence) Construct validity (PDI, SF-36, VAS)	±/±/- ±/±/-	Gross[26], 2008
Isernhagen Work System (IWS)	N = 321 D = back pain T = mean 450-737 days	Predictive validity (time receiving benefit; time to claim closure; recurrence) Construct validity (PDI, VAS)	±/-/- ±	Gross[27-31], 2003, 2004 (2x), 2005, 2006
	N = 336 D = upper extremity disorder T = mean 468 days	Predictive validity (time receiving benefit; time to claim closure; recurrence)	±/±/-	Gross[32], 2006
Multiperspective Multidimensional Pain Assessment Protocol (MMPAP)	N = 599 D = chronic pain T > 180 days	Predictive validity (employment after 8 months)	+	Rucker[33], 1995

N = number of patients; D = disorder; T = days from injury to test
Outcome: - = poor; ± = moderate; + = good
NRS = Numeric Rating Scale; PDI = Pain Disability Index; SF = Short-Form; VAS = Visual Analogue Scale for pain

range from zero (no disability) to 24 (severe disability). The time to administer the questionnaire is 5-10 minutes. In claimants for workers' compensation, the RDQ showed good internal consistency (Cronbach's alpha 0.95), no floor and ceiling defects, and moderate to large responsiveness (response mean 0.78 – 0.84 for improvement). Construct validity was good with positive correlations with the Numeric Rating Scale (NRS-101) and the Short Form 12 and 36 (SF 12 and SF-36) (r = 0.70 – 0.85). Predictive validity was good with positive prediction of return to work outcomes[23;24].

Table 3. Description of the instruments

	Way of assessment	Time to complete	Number of items	Item scale	Target population
RMDQ	Questionnaire	5-10 minutes	24	dichotomous (yes/no)	low back pain
PSFS	Questionnaire	5 minutes	3-5	10 point	musculoskeletal problems
IWS	Performance test	5 hours in 2 days	28	numeric, ratio	musculoskeletal problems
MMPAP	2 Physicians + Questionnaire	2-4 hours	65	3-6 point	chronic pain

RMDQ, Roland-Morris Disability Questionnaire; PSFS, Patient-Specific Functional Scale; IWS, Isernhagen Work System; MMPAP, Multiperspective Multidimensional Pain Assessment Protocol.

Patient-Specific Functional Scale (PSFS)

The PSFS[34] is a patient-specific outcome measure, which was designed for use in patients with varied musculoskeletal problems. It measures functional status by asking the patient to name up to five activities which they are having difficulty performing because of their problem. After specifying the activities, patients are asked to rate each activity on a 0 to 10 scale, with 0 representing the inability to perform the activity and 10 the ability to perform as well as before the onset of symptoms. Time to administer is about 5 minutes. In workers' compensation claimants the PSFS showed moderate construct validity with correlations of 0.32 tot 0.53 for the Pain Disability Index (PDI), 0.32 to 0.44 for the SF-36 and 0.19 tot 0.24 for the Visual Analogue Scale for pain (VAS). As for predictive validity, the association between functional scales and suspension of benefits, an adjusted hazard ratio (HR) of 1.16 (95% CI, 1.07 -1.27) was found, and for claim closure an HR of 1.14 (95% CI, 1.06 – 1.22). No significant association with recurrence of claiming was found[25].

Iser*nhagen Work System (IWS)*

Functional capacity evaluations (FCE) are standardized batteries of clinical tests that measure the performance of the patient or worker, such as lifting, carrying, pushing and walking. There are several FCE protocols, and the IWS is one of them. The IWS consists of 28 tests that reflect work-related activities. During administration of the IWS, the clinician relies on observation of biomechanical and physiological signs of effort to determine safe, maximal performance levels[35]. It is a two-day test, taking 2-3 hours each day, with some tests being repeated on day two. Patients' performance on the test can be matched to the specific tasks of jobs. In claimants for workers' compensation, the IWS showed moderate construct validity in correlation with the PDI (r = 0.44 – 0.55) and the VAS (r = 0.34 – 0.45). As for predictive validity, better lift performance was associated with a shorter time receiving benefit (for upper extremity disorder HR 1.55; 95% CI, 1.29-1.87; for low back pain HR 1.48; 95% CI, 1.14 – 1.92) and time to claim closure (for upper extremity disorder HR 1.81; 95% CI, 1.49 – 2.20; for low back pain HR 1.17; 95% CI, 0.91 – 1.50)[26-31]. There was no association[31] or even a negative association[28] with recurrence.

Multiperspective Multidimensional Pain Assessment Protocol (MMPAP)

The MMPAP is a pain assessment tool that collects and uses information from patient self-reports and medical examination, along with an assessment by two physicians separately. Major domains assessed in the protocol include pain, mental health status, social support, medical information, functional limitations, and abilities and rehabilitation potential[32]. The patient's assessment of the functional limitations includes 17 items in the ADL domain (e.g., climbing stairs and travelling); the physician's rating of the functional abilities domain also includes 17 items (e.g., standing, sitting and lifting). The length of time to complete the MMPAP was patient specific and varied from two to four hours. In workers' compensation claims, the MMPAP showed a good predictive validity (90% for outcome employment)[32].

DISCUSSION

We systematically reviewed the literature on instruments that assess functional limitations in workers' compensation claimants and found studies on four instruments: two questionnaires (RMDQ, PSFS), a performance test (IWS) and an instrument which combined a questionnaire and examination by physicians (MMPAP). For all of these four instruments, the predictive validity was assessed, for three of them the construct validity, and for none of them the reliability. The predictive validity was good in the RDQ and MMPAP, and ranged from poor to moderate in the IWS and PSFS. The construct validity was good in the RDQ, poor to moderate in the IWS and PSF, and not measured in the MMPAP. The instruments assessed a range varying between three to 34 physical functional limitations. No instruments were found for assessing mental functional limitations for claimants, even though psychological complaints are responsible for 35% of the claims for workers' compensation in the Netherlands[36].

This is the first review focusing on instruments for assessing functional limitations in claimants for disability benefits. Because we reviewed the literature systematically in four major electronic databases and checked for additional literature in references, we assume we have included all relevant instruments. However, we may have missed some instruments that have not been published in peer-reviewed journals.

In clinical and rehabilitation settings, the psychometric properties of the four instruments we found have been investigated more frequently. In a rehabilitation setting, the RDQ[15;37] and the PSFS[34;38-40] have been demonstrated as valid, reliable and responsive to change in various conditions. As for the IWS, in a rehabilitation setting, inter-rater reliability and predictive validity were good; concurrent validity was low to moderate[19].The MMPAP proved to be a reliable and valid tool in a population of 651 patients[41]. Our findings on the psychometric qualities of RMDQ and IWS in claimants were in line with these studies, although in most of the studies the instruments were used to evaluate the results of therapy in a clinical setting and not as an instrument to assess functional limitations in claimants. Assessing the validity of instruments measuring functional limitations is a problem because there is

no gold standard, and often it is unclear if it is performance or capacity that has to be assessed. For instance, does a patient with non-specific low back pain have a reduced lifting capacity? And if so, how many kilograms is the patient able to lift? In the studies identified in this review the predictive validity was assessed by measuring "return to work" or the time during which the claimant did receive a benefit. However, these measures may underestimate functional limitations, for instance, when a patient with functional limitations changes functional job status, resumes work part-time or cannot find a job[42]. They may also overestimate the patient's functional limitations, for instance, if a patient is not motivated to go to work. Therefore, "return to work" and "time receiving benefit" are not well suited for measuring validity in instruments that measure functional limitations.

We found two questionnaires that measured functional limitations in claimants. Looking at the content of both questionnaires, there are several issues which make them less suitable for assessing functional limitations in workers' compensation claimants. The questionnaires are not work oriented, but mainly measure limitations in daily life, which makes them more suitable for a rehabilitation setting. There is a mixed content: not only functional limitations are measured but also activity limitations and participation restrictions. Only a few functional limitations are assessed and in the RDQ there is no grading of the functional limitations. Furthermore, there is only a registration of the claimant's perceived limitations, but no real assessment is being conducted. For instance, if a patient claims he cannot walk at all, no assessment is made if this is reasonable considering the patient's disease.

The IWS, on the other hand, is work oriented, measures 28 physical items and gives a grading of these items. It can be used for all somatic disorders instead of a specific disorder such as low back pain. Unfortunately, the validity in the studies we found was poor to moderate and no studies on reliability for claimants could be found. The IWS measures patients' performance; in addition, there has to be an assessment of the sincerity of the patient's effort, the ability to perform work outside a laboratory setting, and whether activities are considered medically safe[43]. One disadvantage is the fact that it takes two days and two to three hours each day to execute the IWS. An abbreviated IWS may offer an efficient alternative[30].

Questionnaires and performance tests are designed to assess patients' actual functioning in a given situation. In claim assessment, however, the ability to perform is the issue that has to be assessed. For instance, if a patient does not perform a certain task, this can be due to a motivational problem or inadequate behavior. In daily practice, physicians often play an important role in this assessment[44]. Therefore, it is remarkable that almost no studies are conducted into the reliability and validity of physicians' assessments. We only found one study in the year 1995[32] that described a protocol for physicians, the MMPAP. Although research into claimants was limited to predictive validity, in a clinical setting satisfactory reliability and validity was found.

CONCLUSION

To summarize, in the present review study, only four instruments were found for assessing the functional limitations in claimants for workers' compensation. Of these four instruments found, the psychometric qualities were not satisfactorily demonstrated. Because the assessment of functional limitations in employees applying for a workers' compensation benefit not only has immense individual implications, but also implications for society as a whole, more evidence-based instruments need to be developed in future research. Performance tests and questionnaires alone cannot properly assess functional limitations without an appraisal of the outcome of these tests. Specialized physicians together with instruments such as performance tests and questionnaires looks the most promising. Further research into the validity of instruments would require a gold standard. Maybe this could be approached by seeking consensus among a number of physicians after a medical examination, an interview protocol, questionnaires and performance tests.

Reference List

1. Abdel-Moty E, Fishbain DA, Khalil TM, Sadek S, Cutler R, Rosomoff RS et al. Functional capacity and residual functional capacity and their utility in measuring work capacity. Clin J Pain 1993;9:168-173.
2. Velozo CA. Work evaluations: critique of the state of the art of functional assessment of work. The American Journal Of Occupational Therapy 1993;47:203-209.
3. Krause N, Frank JW, Dasinger LK, Sullivan TJ, Sinclair SJ. Determinants of duration of disability and return-to-work after work-related injury and illness: challenges for future research. Am J Ind Med 2001;40:464-484.
4. International classification of functioning, disability and health. Geneva: World Health Organization, 2001.
5. Innes E, Straker L. Validity of work-related assessments. Work 1999;13:125-152.
6. Innes E, Straker L. Reliability of work-related assessments. Work 1999;13:107-124.
7. Schultz IZ, Crook J, Berkowitz J, Milner R, Meloche GR. Predicting return to work after low back injury using the Psychosocial Risk for Occupational Disability Instrument: a validation study. J Occup Rehabil 2005;15:365-376.
8. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. 2 ed. Upper Saddle River, New Jersey, Prentice Hall, Inc, 2000.
9. Menard MR, Hoens AM. Objective evaluation of functional capacity: medical, occupational, and legal settings. J Orthop Sports Phys Ther 1994;19:249-260.
10. Kreider B. Latent work disability and reporting bias. J Hum Resour 1999;34:734-769.
11. Spanjer J. De reproduceerbaarheid van WAO-beoordelingen, een literatuuronderzoek. [Reproducibility in disability assessment: a literature search]. Tijdschrift voor bedrijfs- en verzekeringsgeneeskunde 2001;9:195-8.
12. Matheson LN, Kaskutas V, McCowan S, Shaw H, Webb C. Development of a database of functional assessment measures related to work disability. J Occup Rehabil 2001;11:177-199.
13. Flores L, Gatchel RJ, Polatin PB. Objectification of functional improvement after nonoperative care. Spine 1997;22:1622-1633.
14. Innes E. Reliability and Validity of Functional Capacity Evaluations: An Update. International Journal of Disability Management Research 2006;1:135-148.
15. 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.
16. Williams RM, Schmuck G, Allwood S, Sanchez M, Shea R, Wark G. Psychometric evaluation of health-related work outcome measures for musculoskeletal disorders: a systematic review. J Occup Rehabil 2007;17:504-521.
17. Söderberg E, Alexanderson K. Sickness certification practices of physicians: a review of the literature. Scan J Public Health 2003;31:460-474.
18. King PM, Tuckwell N, Barrett TE. A critical review of functional capacity evaluations. Phys Ther 1998;78:852-866.
19. Gouttebarge V, Wind H, Kuijer PP, Frings-Dresen MH. Reliability and validity of Functional Capacity Evaluation methods: a systematic review with reference to Blankenship system, Ergos work simulator, Ergo-Kit and Isernhagen work system. Int Arch Occup Environ Health 2004;77:527-537.
20. Wind H, Gouttebarge V, Kuijer PP, Frings-Dresen MHW. Assessment of functional capacity of the

musculoskeletal system in the context of work, daily living, and sport: a systematic review. J Occup Rehabil 2005;15:253-272.

21. Mayer J, Mooney V, Matheson L, Leggett S, Verna J, Balourdas G et al. Reliability and validity of a new computer-administered pictorial activity and task sort. J Occup Rehabil 2005;15:203-213.
22. Terwee CB, Bot SDM, de Boer MR, van der Windt DIAWM, Knol DL, Dekker J et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34-42.
23. Turner JA, Fulton-Kehoe D, Franklin G, Wickizer TM, Wu R. Comparison of the Roland-Morris Disability Questionnaire and generic health status measures: a population-based study of workers' compensation back injury claimants. Spine 2003;28:1061-1067.
24. Baldwin ML, Butler RJ, Johnson WG, Cote P. Self-reported severity measures as predictors of return-to-work outcomes in occupational back pain. J Occup Rehabil 2007;17:683-700.
25. Gross DP, Battie MC, Asante AK. The Patient-Specific Functional Scale: validity in workers' compensation claimants. Arch Phys Med Rehabil 2008;89:1294-1299.
26. Gross DP, Battie MC. Construct validity of a kinesiophysical functional capacity evaluation administered within a worker's compensation environment. J Occup Rehabil 2003;13:287-295.
27. Gross DP, Battie MC, Cassidy JD. The prognostic value of functional capacity evaluation in patients with chronic low back pain: part 1: timely return to work. Spine 2004;29:914-919.
28. Gross DP, Battie MC. The prognostic value of functional capacity evaluation in patients with chronic low back pain: part 2: sustained recovery. Spine 2004;29:920-924.
29. Gross DP, Battie MC. Factors influencing results of functional capacity evaluations in workers' compensation claimants with low back pain. Phys Ther 2005;85:315-322.
30. Gross DP, Battie MC, Asante A. Development and validation of a short-form functional capacity evaluation for use in claimants with low back disorders. J Occup Rehabil 2006;16:53-62.
31. Gross DP, Battie MC. Does functional capacity evaluation predict recovery in workers' compensation claimants with upper extremity disorders? Occup Environ Med 2006;63:404-410.
32. Rucker KS, Metzler HM. Predicting subsequent employment status of SSA disability applicants with chronic pain. Clin J Pain 1995;11:22-35.
33. Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. Spine 1983;8:141-144.
34. Stratford P, Gill C, Westaway M, Binkley J. Assessing disability and change on individual patients: a report of a patient specific measure. Physiother Can 1995;47:258-263.
35. Isernhagen SJ, Hart DL, Matheson LM. Reliability of independent observer judgments of level of lift effort in a kinesiophysical Functional Capacity Evaluation. Work 1999;12:145.
36. http://statline.cbs.nl . 2009.
37. Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. Physical Therapy 2002;82:8-24.
38. Westaway MD, Stratford PW, Binkley JM. The patient-specific functional scale: validation of its use in persons with neck dysfunction. J Orthop Sports Phys Ther 1998;27:331-338.
39. Cleland JA, Fritz JM, Whitman JM, Palmer JA. The reliability and construct validity of the Neck Disability Index and patient specific functional scale in patients with cervical radiculopathy. Spine 2006;31:598-602.
40. Jolles BM, Buchbinder R, Beaton DE. A study compared nine patient-specific indices for musculoskeletal disorders. J Clin Epidemiol 2005;58:791-801.

41. Rucker KS, Metzler HM, Kregel J. Standardization of chronic pain assessment: a multiperspective approach. Clin J Pain 1996;12:94-110.

42. Fishbain DA, Rosomoff HL, Goldberg M, Cutler R, Abdel-Moty E, Khalil TM et al. The prediction of return to the workplace after Multidisciplinary Pain Center treatment. Clin J Pain 1993;9:3-15.

43. Pransky GS, Dempsey PG. Practical aspects of functional capacity evaluations. J Occup Rehabil 2004;14:217-229.

44. De Boer WEL, Besseling JJM, Willems JHBM. Organisation of disability evaluation in 15 countries. Pratiques et Organisations de Soins 2007;38:205-217.

CHAPTER 4

Disability assessment interview: the role of detailed information on functioning in addition to medical history-taking

Jerry Spanjer
Boudien Krol
Roel Popping
Johan W. Groothoff
Sandra Brouwer

Published in: Journal of Rehabilitation Medicine 2009;41:267-72.

ABSTRACT

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 nine 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 forms 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 about 80%). The physicians who received either medical information or both forms of information indicated fewer work limitations compared to 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 to 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.

INTRODUCTION

The assessment of the work limitations of a patient 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 objectively assess their own work limitations[5]. Furthermore, the patient has a financial interest and 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 two years if an employee is unable to work due to disability. After these two 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 two 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 inquires after, 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 inquire after 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 possibly is a 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 inquire thoroughly after 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 for 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 nine Dutch social insurance physicians were asked to assess patients' work limitations and record them in the Functional Ability List (FAL)[10] by presenting them 30 written patient reports. All 27 physicians had to assess the same patients but each group of nine 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 we obtained a total of 810 patient assessments. 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 indicated 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 those, 26 physicians were willing to cooperate, three were not able to cooperate due to long-term absence, and one physician 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, which is a response rate of 90%. The average length of time spent by these physicians in professional practice was 13 years (range 5–31).

Patients

Thirty patients working in health-care 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 filling the items of the FAL. Patients with these diagnoses represent about 30% of the entire population applying for a social disability benefit. Half of the remaining population apply because of mental

problems and the other half have problems such as neck and upper extremity complaints, heart and lung diseases or carcinoma. The selected patients were diagnosed as follows: twelve patients with low back problems (spinal fracture, herniated disc, M. Scheuerman, chronic non-specific low back pain), six patients with fibromyalgia, four patients with knee problems, two patients with hip problems, two patients with rheumatoid arthritis and four patients with generalized arthrosis. The mean age of the patients was 48.9 years (range 30–63) and 80% were women. The mean duration of sick leave was 3.3 years (range 1–10). The patients had worked on average for 10.3 years at their last job (range 1–27) for 23.2 hours a week (range 2–48).

Reports

A written report consisted of 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 minutes 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 AM, washed, got dressed and drank a cup of coffee. At 6.15 AM I went to work. Then I loaded the car with about 60 crates, each with eight loaves of bread. At 7.30 AM I drove off and went to two shops to deliver the bread. At 8.45 AM 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 PM 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 PM.

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 four-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 (one litre 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 very 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 more than 90% is considered excellent[23].

The Mann-Whitney 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 1 presents the average percentages of linear weighted agreement within the three 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 to the physicians who were either given the medical and the complete versions. Significant differences were found between those

Table 1. 'Linear weighted percentage agreement' between the physicians within the three versions of reports (column 2-4) and significant differences between the three versions on scores of the Functional Ability List items (column 5-7)

Items	Percentage agreement			Significant differences		
	Medical	Functional	Complete	FxM	FxC	MxC
Body movement scale						
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 ↑
Body posture scale						
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**	**n=13**	**12**	**7**

* = dichotomous data, other items are ordinal; ↑ = more serious limitations
F = functional version; M = medical version; C = complete version

who received the medical and the complete versions in 7 out of 21 items. Those using the medical version revealed more serious limitations three times, and the physicians using the complete version did so four times.

Table 2 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

Table 2. Number of hours a week each patient (Pat) could work in each of the three versions according to the physicians

	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**
Pat 1/12																				
Pat 13		30							20			30	20	20		20				
Pat 14				20																
Pat 15															30					30
Pat 16																30				
Pat 17							20													
Pat 18															20					
Pat 19				30	30								30		30			30		30
Pat 20				30	30															
Pat 21				30								30								
Pat 22				30																
Pat 23																10				
Pat 24				30								30		20						
Pat 25			30	30	20	20	20	20	20		20	20		20		20			30	
Pat 26				20								30								
Pat 27				30																
Pat 28	20	30			20		20	20	20				30	20		30			30	30
Pat 29												30		30		30				
Pat 30				30																

Empty cell = full-time

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 3 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 three versions of the reports.

Table 3. Information provided and percentage of additional information needed according to the insurance physicians (n = 3x9) in each of the three versions (n = 3x30)

Items	Complete		Medical		Functional	
	Provided	Infoneed %	Provided	Infoneed %	Provided	Infoneed %
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)
Infoneed = 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 were only provided with medical information mainly indicated a need for more information 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 three groups indicated a need for additional information from the treating physician in 21 to 34% of the cases.

DISCUSSION

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

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 to 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 to 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 four times less often (6 times versus 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 to 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 inquiring after detailed examples), alongside medical history, in disability assessment.

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

In this study the assessments were based on written reports for practical reasons. 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 only provided 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.

CONCLUSION

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.

Reference List

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;30:130-40.
2. Pransky GS, Dempsey PG. Practical aspects of functional capacity evaluations. J Occup Rehab 2004;14:217-29.
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;85:837-9.
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. Disability and Rehabilitation 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;61:1447-52.
6. Menard MR, Hoens AM. Objective evaluation of functional capacity: medical, occupational and legal settings. J Orthop Sports Phys Ther 1994;19:249-60.
7. Boot CRL. 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, Ende CH. van den. Impairment measures in rheumatic disorders for rehabilitation medicine and allied healthcare: a systematic review. Reumatol Int 2005;25:501-12.
9. Boer WEL de, Besseling JJM, Willems JHBM. Organisation of disability evaluation in 15 countries. Pratiques et Organisations de Soins 2007;38:205-17.
10. ClaimBeoordelings- en BorgingsSysteem (CBBS), Lisv Amsterdam; 2002.
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]. CTSV Zoetermeer; 2000.
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;11:360-7.
13. Spanjer J. De reproduceerbaarheid van WAO-beoordelingen, een literatuuronderzoek. [Reproducibility in disability assessment: a literature search]. Tijdschrift voor bedrijfs- en verzekeringsgeneeskunde 2001;9:195-8.
14. International classification of functioning, disability and health. WHO: Geneva; 2001.
15. Köhnken G. Statement validity analysis and the detection of the truth. The detection of deception in forensic contexts. Granhag, Pär-Anders, Strömwall, Leif. NY 2004; 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 and criminological psychology 2005;10:225-45.
17. Tredoux C, Pooley S. Credibility Assessment. In: Tredouw C, Foster D, Allan A, Cohen A, Wassenaar D. 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. Psychology, Public Policy and 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;10:162-3).

20. Popping, R. On agreement indices for nominal data. In: Saris, WE & Gallhofer IN (eds.), Sociometric Research, Volume 1, Data Collection and Scaling. London: MacMillan Press, 1988; 90-105.
21. Popping, R. AGREE, A package for computing nominal scale agreement. Computational Statistics and Data Analysis 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 1972;7:17-18.
23. Landis JR, Koch GG. The measurements of observer agreement for categorical data. Biometrics 1977;33:159-174.

CHAPTER 5

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.

An abstract was published in: Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2005;13:10.

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)[1]. The insurance physician registers in this list which work limitations the patient has and to what degree. The list contains 28 items of which 27 are physical and 1 is mental. 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[2]. 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)[3]. 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[4]

- 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 testing[7]. 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 disability[8-10]. Often a wide inter-rater variability between physicians is present when assessing work limitations[11,12].

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)[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

on the 6 patients with mental complaints.

The statistical software package AGREE 7.0[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[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)

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

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

Physician	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**N**
Patient													
1	5x4	5x4		5x4				5x6	5x6		5x4		**6**
2			5x5					5x6	5x4				**3**
3			5x6										**1**
4													**0**
5	5x4												**1**
6	FI	5x4	5x4			FI		FI			5x4		**6**
7			5x4					5x4					**2**
8													**0**
9	FI	FI	FI		FI	FI	FI	12	5x1	FI	FI		**10**
10													**0**
11	3x4		3x4	5x4			3x4				5x4		**5**
12	5x4	5x4	5x4					10		FI			**5**
N	**6**	**4**	**7**	**2**	**1**	**2**	**2**	**6**	**3**	**2**	**4**	**0**	

FI = 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[5]. 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[5]. 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'[18] 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)[19,20]. In general, most physical work-related instruments are based on the DOT[21]. 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[3] 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)[22]. 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[6], 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

1. Functie Information Systeem (FIS). Naslag gevalsbehandeling [Instruction social disability benefit assessment]. Houten/Diegem: Bohn, Stafleu, van Loghum, 1996.
2. Boer WEL de, Wijers JHL, Spanjer J, Beijl I van der, Zuidam W, Venema A. Gespreksmodellen in de verzekeringsgeneeskunde. [Interview models in insurance medicine]. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2006;14:17-23.
3. Spanjer J. Belastbaarheid gerichte beoordeling, beschrijving van een beoordelingsmethodiek. [Disability Assessment Structured Interview (DASI), description of an assessment method] 2004. A copy can be acquired from the author.
4. 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. (Reactie lezersforum 2002:10;162-3).
5. Spanjer J. De inter- en intra-beoordelaarsbetrouwbaarheid van WAO beoordelingen. [The inter- and intra-rater reliability of disability pension assessments]. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2001;9:234-241.
6. Brouwer S, Dijkstra PU, Gerrits EHJ, Schellekens JMH, Groothoff JW, Geertzen JHB, Göeken LNH. Intra- en inter-beoordelaarsbetrouwbaarhied 'FIS-belastbaarheidspatroon' en 'Functionele Mogelijkhedenlijst'. [Intra- and inter-rater reliability 'Functional Information System' and 'Functional Ability List']. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2002;10;360-367.
7. Brouwer S, Dijkstra PU, Stewart RE, Göeken LHN, 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 Rehab 2005;27:999-1005.
8. Loeser JD, Sullivan M. Docters, Diagnosis and Disability. Clinical Orthopaedics and related research 1997;336:61-6.
9. Ehrlich GE, Wolfe F. On the difficulties of disability and its determination. Rheumatic Diseases Clinics Of North America 1996;22 ;613-21.
10. Tait RC, Chibnall JT, Andersen EM, Hadler NM. Disability determination: validity with occupational low back pain. The Journal of Pain 2006;7:951-7.
11. Innes E, Straker L. Reliability of work-related assessments. Work 1999;13:107-24.
12. Wolfson AM, Doctor JN, Burns SP. Clinician judgments of functional outcomes: how bias and perceived accuracy affect rating. Arch Phys Med Rehabil 2000;81:1567-74.
13. Spanjer J. De praktische concretisering van psychische belastbaarheid. [The Mental Ability List]. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 1996;5:182-6.
14. Popping R. More agreeable software? Computational Statistics & Data Analysis Statistical Software Newsletter 1992;13: 223-7.
15. Schouten HJA Measuring pairwise agreement among many observers. Biometrical Journal 1980;22: 497-504.
16. Schouten HJA Measuring pairwise agreement among many observers. II. Some improvements and additions. Biometrical Journal 1982;24: 431-5.
17. Landis JR, Koch GG. The measurements of observer agreement for categorical data. Biometrics 1977;33:159-174.
18. Standaard verminderde arbeidsduur. [Guideline reduced working hours]. Amsterdam: Lisv, 2007.

19. U.S Department of Labor, Employment and Training Administration. Dictionary of occupational titles. Fourth Edition. Supplement. Washington, DC: U.S. Government Printing Office, 1986.

20. Fishbain DA, Cutler RB, Rosomoff H, Khalil T, Abdel-Moty E, Steele-Rosomoff R. Validity of the dictionary of occupational titles residual functional capacity battery. Clin J Pain 1999;15:102-10.

21. Innes E, Straker L. Validity of work-related assessments. Work 1999;13:125-152.

22. International classification of functioning, disability and health. Geneva: World Health Organisation, 2001.

CHAPTER 6

Inter- and intra-rater reliability in disability assessment based on "Disability Assessment Structured Interview" (DASI) video recordings

Jerry Spanjer

This chapter is a translation of: De inter- en intra- beoordelaars betrouwbaarheid van WAO beoordelingen. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2001;9:234-41.
Reprinted with the kind permission of TBV.

ABSTRACT

Objective

Insurance physicians are expected to conduct reproducible assessments of the functional limitations for work disability pensions. The "Disability Assessment Structured Interview" (DASI) is a semi-structured interview for assessing functional limitations in work disability claimants.

The aim of this study was to analyze the content of DASI interviews, ask physicians for comments and determine the inter- and intra-rater reliability of assessments made using the DASI

Methods

Five insurance physicians, using the DASI, made 14 video recordings of first-time disability assessment interviews. The recordings were analyzed by measuring the duration of the different topics of the interviews.

The video recordings were shown to 22 insurance physicians who were asked for their comments on the interviews. The inter-rater reliability among the 22 physicians was measured by computing the percentage agreement with respect to the mental and physical items of two Dutch disability lists, the Functional Information System (FIS) and the Mental Ability List (MAL).

To measure the intra-rater reliability, the five insurance physicians who made the recordings were asked to fill out the FIS and MAL right after the recordings were made and again after seeing the video after six months.

Results

The mean duration of the interviews was 33 minutes (range 19-77 min.), and as much time on impairments was spent as on functional limitations and activities (each comprising 30% of the whole interview). In general, the 22 physicians found that the interviews were structured, functional and efficient, and that in their own interviews more attention was given to medical issues and less to the functional limitations experienced by the patients. The mean percentage agreement among the 22 physicians on the items of the FIS and MAL was 74% (range 56-85%). There was a considerable difference among the physicians in assessing the hours a patient can work daily. The intra-rater percentage agreement was 80% (range 52-100%).

Conclusion

The DASI interviews took 33 minutes on average, and as much time was spent on impairments as on functional limitations and activities. In general, the physicians found that the interviews were structured, functional and efficient. The inter- and intra-rater reliability on most mental as well as physical items was moderate to good. The inter-rater agreement on "the hours a patient can function daily" was low.

INTRODUCTION

In the "Assessment Decree Occupational Disability Act"[1] social insurance physicians are expected to conduct reproducible assessments of functional disability. This means that it should be likely for fellow physicians to arrive at the same findings and conclusion. In statistics, the concept of reliability is used: inter-rater reliability for the agreement in assessments among different raters, and intra-rater reliability for the agreement between assessments by the same rater.

The sparse literature on research into the reproducibility of functional disability assessments indicates a low agreement among insurance physicians[2]. The sources for this low agreement lie in three factors: physicians start from different basic assumptions in making their assessments, the information on which the assessment is based (mainly the patient interview) is not unambiguous, and the interpretation of the collected information (using protocols and guidelines) among physicians differs. The literature also indicates that physicians mainly base their assessments on concrete activities patients undertake, while often not paying attention to the detailed and concrete activities in their interviews[3,4]. Therefore, this study uses a semi-structured interview, the Disability Assessment Structured Interview (DASI), which specifically pays attention to the concrete and detailed examples of functional limitations and activities.

The aim of this study is to evaluate the reproducibility of the assessment of functional limitations in the claims of work disability claimants when a semi-structured interview is used, and to analyze the interviews themselves.

The research questions are:

1. Which topics are addressed in the DASI interviews in daily practice and how much time is spent on them?
2. What comments do insurance physicians make about DASI interviews?
3. What is the inter- and intra-rater reliability of assessments that use the DASI?

METHODS

DASI interviews of patients conducted by insurance physicians were recorded on video. The recordings were analyzed by measuring the duration of the different topics of the interviews. The video recordings were then later shown to insurance physicians, who assessed the functional limitations of the patients shown and made comments about the interviews.

The reliability of the assessments is, among other things, found to be influenced by[5]:

- The definition of the assessed items
- The assessed patients
- The assessment method
- The setting and period of the assessment

- The assessors.

The definition of the assessed items

In quantifying the assessed functional limitations, the Function Information System List (FIS)[6] was used. The physical part of this list consists of 27 items with a two- to ten-point rating scale. In the FIS the assessors can also indicate whether patients are limited in the amount of hours they can work daily, and whether patients are unable to work on evening or night duty or in irregular shifts. The insurance physicians use the FIS every day in their regular work; therefore, the definitions of the items were already known to the physicians.

Because mental limitations are not readily quantifiable in the FIS and because physicians are not able to handle "Item 28" of the FIS very well[7], the Mental Ability List (MAL) as described by Spanjer[8] was used to record the mental limitations. In the MAL, eight items are described with a three to five-point rating scale.

The assessed patients

Variation in assessments depends on the patients; in patients with an unambiguous specific functional limitation (for instance, deafness), less variation in assessment outcome can be expected than in patients with more diffuse problems (for instance, chronic fatigue syndrome). Because of this, a certain selection of patients needed to take place. Patients applying for a first-time disability benefit after one year of sick leave (WAO) were selected for the following items:

- The DETAM as well as BVG population of the Social Security Office in Groningen, the Netherlands.
- The first day of the invalidity benefit (one year after sick listing) was in April 2000.
- Not very complicated or very simple problems.
- 50% patients with mental complaints and 50% with physical complaints.
- Not obviously either fully able or unable to work.

Fifteen patients agreed that the interview could be recorded on video. Of these, seven patients had a mental diagnosis (overstrained, depression, burnout) and seven a physical diagnosis (shoulder complaints, low back pain, whiplash, sinus problems, herniated muscle, broken leg, thyroid gland problems); one patient cancelled his appointment at the last moment. Three patients had physical as well as mental complaints.

The assessment method

To assess the functional limitations, we used the Disability Assessment Structured Interview (DASI)[9]. The DASI is a semi-structured interview that focuses a great deal of attention on specific and detailed examples of limitations and concrete activities which the patient still undertakes. In Box 1, the specific topics of the DASI are described. The DASI was chosen for practical reasons: the physicians

who recorded the interviews had all been trained in administering the DASI interview. Furthermore, a relatively high agreement in assessed functional limitations among the assessing physicians was expected, because in DASI interviews a lot of attention is spend on functional limitations. Moreover, because the DASI has a simple structure, it was relatively simple to comment upon several topics of the interview.

Box 1. Topics of the "Disability Assessment Structured Interview" (DASI)

Introduction
- putting the patient at ease
- short explanation of the procedure
- summarizing the known data.

Work
- type and duration of the work
- content of the occupation
- perception of the work by the patient.

Impairments
- medical history and nature of current complaints
- other diseases and general anamnesis
- course of the disease
- cause of the disease (the treating physician's diagnosis; private/work stress; personality)
- treatment and medication use (now and in the past)
- if necessary, more information from the treating physician, hetero-anamnesis, expert consultation with a specialist.

Activity limitations
- limitations experienced in daily life and at work, for example, lifting, walking and bending.
- concrete and detailed examples of the limitations experienced.

Participation
- activities of daily living (ADL), description of a normal day, hobbies, sports, housekeeping, social contacts, and work. The patient is asked which activities are actually performed and for how long.

Patient's opinion
- patient's view of his or her work limitations
- patient's response to the provisional opinion of the physician.

Physician's opinion
- the physician communicates his or her judgment
- explanation of the rest of the procedure
- opportunity for the patient to respond.

Setting and period of assessment

It would be ideal if different assessing physicians could assess the same patient at the same time individually. If one patient is assessed by different physicians at different points in time, subsequent assessments can be influenced by previous assessments. Furthermore, undergoing several interviews is too great a burden on the patients. Since in written reports information can

go lost because the physicians do not see the patient, in this study we chose video recordings of assessment interviews.

Five insurance physicians employed at Cadans, Groningen (the Netherlands), agreed to cooperate in this study. One of the physicians (10 years of experience in disability assessments) was the instructor of four other physicians (one to two years of experience in disability assessments). The four physicians were trained to administer the DASI in daily practice, and had an additional instruction session for half a day.

In January 2000, the five physicians each assessed two to three patients using the DASI. A total of 14 interviews were recorded on video. The recordings showed only the patients, while both the physicians as well as the patients could be heard. The physical examination was not recorded on video, but a description of the physical examination was handed out to the assessing physicians.

The assessors

Staff physicians of all social security offices in the Netherlands were asked to supply the names of insurance physicians (experienced as well as inexperienced) who were willing to cooperate in this study. A group of 22 insurance physicians was selected. Selection criteria were: being able to be present on the days the study was conducted, distribution over different social security offices and locations, and registered and not yet registered insurance physicians.

Of the 22 assessing physicians, 12 were employed at GAK, six at Cadans, three at SFB, and one at GUO. Ten physicians were officially registered as insurance physicians, and seven were female. The mean years of experience in disability benefit assessments (WAO) was seven (range 1-25 years).

To measure intra-rater reliability, the physicians who made the recordings of their interviews, looked at their own two or three recordings again after six months, and once again filled out the FIS and MAL.

The 22 assessing physicians were invited for two consecutive days and looked at seven video recordings a day. At the end of each recording the FIS and MAL were filled out, and comments about the interviews conducted were made by the assessing physicians. At the end of each day there was a further evaluation by filling out a questionnaire which asked for general comments about the interviews, recommendations for improvement, and differences with the 22 physicians' own interviews.

Analysis

The comments, recommendations and remarks of the assessors were mapped and summarized. The time the different topics took in the recorded interviews was measured with a stopwatch. To measure the dispersion of the assessment outcomes among physicians, the standard deviation on the FIS and MAL items was calculated. Furthermore, the extend values were scored outside the bandwidth; "the mode or a value above it or under it" was investigated.

The agreement percentages among the 22 assessors were measured for each item of the FIS and the

MAL. We used the statistical computer program AGREE 7.0[10]. This program offers the possibility of measuring the "weighted kappa" and "agreement percentage" for several assessors. It is important to "weight" the outcome among the assessors because that way a distinction can be made between the size of the difference between the scores of the assessors. For instance, for the item "sitting," the difference between scores 1 and 5 is bigger than between 4 and 5. The further the values are apart, the less agreement the assessors have.

We studied the agreement among 22 assessors; this agreement was based on comparing all possible pairs of assessors[11,12]. In the output of AGREE, a weighted kappa and "percentage agreement" was calculated. In general, a "percentage agreement" > 80% is excellent, while 60-80% is moderate to good[13].

RESULTS

Analyses of the interviews

The mean duration of the interviews was 33 minutes, with a mean introduction of 2 minutes and 30 seconds. The range was 19 to 38 minutes, with one exception of 77 minutes. Every interview was analyzed; the duration of the different topics of the DASI as described in Box 1 was measured. In Table 1 the mean duration of the interview and the mean percentage of the time of the different topics for each physician are presented.

Table 1. Mean duration of the interview and percentage of the time of the different topics

Physician	Time (min)	Intro	Work	Impairment	Lim/Act	Opinion pat	Judge	Reaction
1	30	1	8	36	26	11	17	5
2	27	11	11	20	27	8	19	6
3	32	10	15	32	30	4	7	1
4	56	8	7	32	27	12	11	0
5	22	13	3	29	34	3	13	9
Mean	**33**	**9%**	**9%**	**30%**	**29%**	**8%**	**13%**	**4%**

Time (min), mean duration of the interview in minutes; Intro, introduction of the physician; Lim/Act, limitations and activities of the patient; Opinion pat, opinion of the patient concerning his functional limitations; Judge, judgment of the physician; Reaction, reaction of the patient.

Comment of the assessors on the DASI interviews

In general the assessing physicians found that the interviews as seen on the video recordings were fairly uniform; all the items of the DASI were attended to. However, the order in which the topics were dealt with and the conversational skills of the physicians on the video recordings were found to vary. A considerable portion of the assessors indicated that in their own interviews they paid more attention to medical issues (medical history, complaints, therapy and diagnosis) (41% of the assessors) and "claim complaint" (27% of the assessors). They indicated that the medical data in

particular were needed in order to examine the claim; (quote) "The medical issues are not sufficiently attended to. The functional limitations the patient claims have to be tested by medical facts."
Compared with the DASI, the assessors spent less attention on functional limitations in their own interviews, in particular on mental limitations. The DASI was considered structured, functional and efficient, but a more extensive interview on medical issues was advised. They pointed out that a social anamnesis (data considering partner and family) and social functioning were missing.

The assessing physicians were asked which items they found most important in a functional limitations assessment interview. They thought the following items were important in increasing degree of importance: the therapy, the patient's opinion, the medical history, the cause of the disease, the work, the course of the disease, participation, complaints, and the activity limitations. Some striking facts were:

- 16 insurance physicians found that "participation" or "activity limitations" were more important than "complaints."
- 14 insurance physicians found that the "opinion of the patient" was one of the four least important topics (out of the 9 topics).
- 17 insurance physicians found that the "therapy" was one of the four least important topics.

The assessors found that they were able to make adequate assessments of the functional limitations based on the DASI recordings; only in three out of 301 assessments was the FIS or MAL not filled out.

In four patients, a major portion of the assessors found that additional information from the treating physician was needed. In one patient, six out of 22 assessors found they needed an additional expert consultation. Physical examination was found to be insufficient in two patients.
Further observations as a result of the comments:

- Several times a FIS was filled out where there was no disease, but only limitations due to personality.
- Several times the assessor indicated that a FIS would not have been filled out in daily practice because:
  - a considerable improvement was expected within three months
  - patients were expected to be fit for their own job at the time the disability benefit started
  - there was no final medical situation
  - the physicians would await further information from treating physicians or an expert consultation by a specialist.
- It was not always clear what to assess in daily practice: the functional limitations at the moment of examination or, as might be expected, at the time the disability benefit would be set to start.
- In the case of limitations of the arm or hand, a great deal of explanation was added in the FIS.

Inter-rater reliability

Measurements of variability

Table 2 presents the mean standard deviation and the number of times that the assessors deviated more than one gradation from the mode. The more gradations an item contains, the greater the chance that there is a deviation from the mode, of course. Because the mental items have a scale with a maximum of five gradations and the physical items up to ten gradations, the chance for a deviation in physical items is higher. The items with the highest deviations were "working above shoulder level" and "sustained bending." Above average deviation was scored on the items "kneeling," "climbing," "frequent bending," "lifting" and "carrying." Of the mental items, "social interaction" deviated above average. No important differences were found between experienced and inexperienced physicians. In physical items, variation was lowest in patients with unambiguous complaints of low back, arm or leg. An above average variation was found in a patient with vague complaints of the sinuses, a patient with fatigue due to thyroid gland problems, one with neck and shoulder complaints, and a patient who had recently had a leg operation.

In the mental items, the lowest variation was found in nearly recovered patients with psychological complaints due to incompatibility between the person and the work. Variation was highest in the patient with vague complaints of the sinuses, one with psychological complaints due to life events, and a patient who had considerable psychological complaints but who was quite active during daytime.

Percentage agreement

Table 2 presents the agreement percentages of the FIS and MAL items. Among the assessors, a moderate to good agreement existed on the items; the mean agreement percentage was 74%; the range was 70-85%, except for the items "working above shoulder level" (64%) and "emotional pressure" (56%). The agreement percentage among registered physicians (76%) was significantly higher than among non-registered physicians (71%). No significant difference in the level of mean scores on the items was present between these groups.

Hour limitations and limitations in work pattern

If the assessing physicians find a patient cannot work a whole day or week, an "hour limitation" can be scored. Table 3 presents the limitations in "the number of hours a patient can work weekly"; the physicians scored for each patient.

Distinct differences among the physicians existed:

- Five assessors did not score one single hour limitation; the other 17 did score an hour limitation one to seven times across the 14 patients.
- If the physicians thought an hour limitation should be present, little agreement existed on

Table 2. Measurements of variation on the items of the FIS and MAL

Item	SD mean	mode ± 1	%agree(inter)	%agree(intra)
Sitting	0.3	0.0	84	100
Standing	0.9	0.8	83	82
Walking	0.9	0.8	81	80
Climbing stairs	0.7	0.3	77	64
Climbing	0.9	1.1	73	52
Kneeling	1.3	1.8	71	80
Sustained bending	1.4	2.5	74	77
Frequent bending	0.9	1.5	78	70
Reaching	0.9	0.9	73	93
Working above shoulder	1.4	2.6	64	68
Lifting	0.9	1.3	75	75
Carrying	1.2	1.6	72	67
Structure	0.7	0.3	77	100
Responsibility	0.6	0.2	73	83
Time pressure	0.5	0.2	85	100
Emotional pressure	0.6	0.0	56	78
Concentration	0.7	0.2	73	94
Environment	0.9	0.7	71	67
Conflict handling	0.8	0.7	70	93
Social interaction	0.8	1.3	71	83
Mean	**0.9**	**0.9**	**74**	**80**

SD mean, mean standard deviation; mode ±1, mean amount of scores outside the area mode plus or minus one value; %agree(inter), percentage agreement among assessors; %agree(intra), percentage agreement within assessor.

the amount of hours the patient should be limited to. In only 53%, an hour limitation of 10, 20 or 30 hours was given, as the guideline recommended as figures for reduced working hours[14]. Often the physicians complied, filling in the amount of hours the patient really did work, or with the amount of hours the patient should normally work.

- In four patients, no hour limitation was scored. In six patients, an hour limitation was scored by four or fewer physicians. In four patients, an hour limitation was scored by six or more physicians.
- The patients without an hour limitation more often had somatic complaints (75%) as compared to patients who did receive an hour limitation (40%).
- Experienced physicians (6 years experience or more in disability assessment) scored an hour limitation twice as often as inexperienced physicians; this was a significant difference ($p = 0.02$).
- Female physicians significantly scored an hour limitation more often than male physicians ($p = 0.04$).

Table 3. Number of hours a week each patient could work according to the physicians

Physician	1-5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Times	Now	Norm
Patient 1																			**0**	0	ft
Patient 2											5					24	15	12	**4**	12-18	24
Patient 3																			**0**	10	30
Patient 4				28			30								28		28		**4**	28	34
Patient 5																20			**1**	0	11
Patient 6				28													28		**2**	28	ft
Patient 7				20			20	30				20				16		30	**6**	16	32
Patient 8																20			**1**	0	10
Patient 9		12	12		12	12	30	20		12	12	20	20	25	20	12	12		**13**	8	18
Patient 10						20													**1**	0	ft
Patient 11				20	20		20	20				16		10	20	10			**8**	0	27
Patient 12		20	9	9	12	20	20	20	20	20					20	16	9		**12**	0	24
Patient 13																			**0**	0	25
Patient 14																			**0**	0	24
Times	**0**	**2**	**2**	**5**	**3**	**3**	**5**	**4**	**1**	**2**	**2**	**3**	**1**	**2**	**4**	**7**	**5**	**2**			

Times, number of times a limitation in the amount of hours is given; Now, the amount of hours the patient is working at the time of assessment; Norm, the amount of hours the patient normally works when not ill; ft, full time; Empty cell = full time

One patient was considered limited in irregular duty or evening- and nightshifts. This patient indicated herself that this was a problem. Five other patients (often with more serious problems) were considered limited on this item (one to four assessors).

Intra-rater reliability

The mean agreement percentage on the items of the FIS and MAL between the first assessment and the assessment on video after six months was 80%; the range was 64-100%, except for the item "climbing" (52%). In the second assessment, based on the physician's own video recordings, an average higher score on the somatic FIS items was present. Therefore, the assessed abilities of the patients were considered greater. In the mental items of the MAL, this phenomenon was not present; the score in the first assessment was about the same as in the second with a small variation which could be a gradation up or down. Dichotomous items of the FIS which were considered limited in the first assessment often were not considered limited in the second assessment. In two clients, an hour limitation of four hours was scored; this was also the case in the second assessment.

DISCUSSION

The assessment interview

The analyses of the DASI interview video recordings show that as much time was spent on medical as on "functional limitations and activities." The comments of the assessing physicians indicated this was not the case in their daily practice; more time was spent on medical issues than on functional limitations and activities. Nevertheless, the assessing physicians said they thought that, in general, information on functioning was more important than information on medical history-taking. There seems to be a difference between what physicians theoretically think is important in the assessment of functional limitations and what is in fact practiced. Furthermore, to explicitly talk about abilities and disabilities of the patient in the interview seems to favor the validity of the assessment, because the abilities and disabilities are what have to be assessed by the physicians.

In general the DASI interviews supplied sufficient information to conduct a proper assessment, according to the physicians. This is also shown by the fact that in only three out of 301 assessments were FIS or MAL unable to be filled out. If information was missing, this mostly had nothing to do with any shortcomings of the interview, but instead with the physical examination being too concise or lacking information from treating physicians or an expert consultation.

Inter-rater reliability

FIS and MAL items

A moderate to good agreement in the scores was found among the physicians for the FIS items. The fact that a reasonable reliability was found in assessments using the DASI does not mean the validity of these assessments is also reasonable. The assessors can agree on the functional limitations of a

patient (reliability), but this does not prove that this assessment is correct (validity).
We did not measure the agreement among physicians on the exact scores of the FIS and MAL, but a weighted agreement in which a certain bandwidth was considered instead. This is in accordance with the concept which lies behind the FIS; it is an instrument for consultation between insurance physician and labor expert, not an exact instrument to record the functional limitations in decimals. In the mental items of the MAL we also found a moderate to good agreement among assessors. Apparently, it is possible to assess the functional limitations in psychological diseases as well as in physical diseases with a similar agreement among assessors.
Experienced insurance physicians showed a higher agreement on the functional limitation items than less experienced physicians. Apparently, education and experience help in reaching a higher reliability among physicians.
I doubt if the inter-rater reliability as measured in this study could be obtained in daily practice, because in this experiment we used a "laboratory situation." In daily practice each physician has a different interview; there is interaction between patient and physician, and the functional abilities often are not recorded in the FIS list. There were different opinions among the physicians as to what to record in the FIS list: the current functional limitations or the limitations as expected at the start of the disability benefit (after one year of sick listing). Furthermore, some physicians filled out a FIS list for patients who were not ill anymore, but who still had limitations as a result of their personality structure, while other physicians didn't and just stated the patient was not ill anymore.

Hour limitations

There was little agreement among physicians in assessing a reduction in "hours a patient can function daily." If a reduction was presumed, there was little agreement on the amount of hours the patient was still able to function. This, in spite of the fact that the guideline "reduced working hours" [14] was released five months before the assessments took place.
One possible explanation for the fact that experienced physicians more often scored an hour limitation is that they are more oriented towards reintegration of the patient back into his work, than on assessment of the functional limitations. This would seem to be so, because the hour limitations the physicians scored often fitted the amount of hours the patient actually worked. Female physicians scored an hour limitation more often than male physicians. Perhaps this was caused by the fact that most of the patients in the video recordings were female (12 of the 14 patients).

Guidelines

This study shows that in actual practice there is no uniform use of guidelines among insurance physicians. Apparently, the "medical disability criterion" guideline has not created consensus among physicians in how to deal with physical limitations without objective medical findings. The "reduced working hours" guideline has not led to a uniform assessment of the daily hours a patient can work. The "no lasting usable abilities" guideline did not supply a uniform procedure concerning when the FIS list had to be filled out. This could be caused by the fact that physicians have a different

interpretation of these guidelines. What argues against this hypothesis, however, is that even matters which are clearly described in the guidelines are not executed. For instance, the hour limitations that are theoretically possible should be assessed, not the amount of hours the patient actually works; or a FIS list should be filled out even if improvement is expected.
Study into the causes of the lack of uniform use of guidelines and trying to change this could improve the reliability of assessments. Possible actions are: adaptation of guidelines on points that cause differences in interpretation, a thorough introduction to and training in the guidelines, surveillance if guidelines are employed, and feedback on physicians' performance.

Intra-rater reliability
The physicians performed a more "strict" assessment (less functional limitations) when they saw their own video again in a video recording. This is possibly caused by the fact that in the first assessment the physician had actual contact with the patient, while in the second assessment a distance had been created because just a video recording of the patient was shown. This is in line with the comments of the assessing physicians who indicated that they probably were stricter in assessments based on a video recording than in their own consulting room (more distance, no need to inform the patient of the result of the assessment). In any case, there was a satisfactory agreement, a mean agreement percentage of 80%, between video and real life assessments.

CONCLUSION

1. The mean DASI interview took about half an hour (physical examination not included). About as much time was spend on medical impairments as on "functional limitations and activities." Comments of the insurance physicians indicated that in usual practice more attention was spent on medical impairments.
2. In general the DASI interviews gave the physicians enough information to assess the functional limitations.
3. The intra- and inter-rater reliability based on the DASI interviews was moderate to good on physical as well as mental items. However, little agreement among physicians was found on scores on the item "hour limitations."
4. There are signs that the agreement among insurance physicians in normal practice is less than what was measured in this study.

For their cooperation, I'd like to thank:
W.E.L. de Boer, senior researcher, TNO arbeid Hoofddorp
J.P.J. Broersen, psychologist/methodologist, SKB Amsterdam
B.J. Kuipers, insurance physician, Cadans Groningen
R. Popping, assistant professor/researcher, Sociology Department, University of Groningen
M.F. Reneman, coordinator Work exploration, Beatrixoord Groningen

E. Soer, student Human movement science, University of Groningen
N.K.A. Tuk, insurance physician, Cadans Groningen
J.G. Voogd, insurance physician, Cadans Groningen
S. Wiersma, insurance physician, Cadans Groningen
The 22 assessing physicians.

Reference List

1. Schattingsbesluit arbeidsongeschiktheidswetten [Assessment Decree Occupational Disability Act]. Den Haag: Ministerie van Sociale Zaken en Werkgelegenheid, 2000.
2. Spanjer J. De reproduceerbaarheid van WAO beoordelingen, een literatuuronderzoek. [The reproducibility in disability assessment: a literature search]. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2001; 9:195-198.
3. Razenberg PPA. Verzekeringsgeneeskundige oordeelsvorming: inzicht in de praktijk. [Judgement in insurance medicine; insight in practice]. University of Amsterdam 1992 (dissertation).
4. 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, the Netherlands: CTSV, 2000.
5. Hawkins, RP, Fabry BD. Applied behavior analysis and interobserver reliability. A commentary on two articles by Birkimer and Brown. Journal of Applied Behavior Analysis 1979;12: 545-52.
6. Function Information System (FIS), foundations. Naslag gevalsbehandeling Lisv. Houten/Diegem: Bohn, Stafleu, van Loghum, 1996.
7. Evaluatie Fis-methodiek. [Evaluation of FIS methods]. KPMG and University of Maastricht, 1997.
8. Spanjer J. De praktische concretisering van psychische belastbaarheid. [Documentation of mental abilities]. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 1996;5:182-6.
9. Spanjer J. Disability Assessment Structured Interview (DASI), description of an assessment method. Zuidlaren, the Netherlands: 2001. A copy can be acquired from the author.
10. Popping R. More agreeable software? Computational Statistics & Data Analysis Statistical Software Newsletter 1992;13: 223-7.
11. Schouten HJA. Measuring pairwise agreement among many observers. Biometrical Journal 1980;22: 497-504.
12. Schouten HJA. Measuring pairwise agreement among many observers. II. Some improvements and additions. Biometrical Journal 1982;24: 431-5.
13. Landis JR, Koch GG. The measurements of observer agreement for categorical data. Biometrics 1977;33:159-174.
14. Standaard verminderde arbeidsduur. [Guideline reduced working hours]. Amsterdam: Lisv, 2000.

CHAPTER 7

Reliability and validity of the Disability Assessment Structured Interview (DASI): a tool for assessing functional limitations in claimants

Jerry Spanjer
Boudien Krol
Sandra Brouwer
Roel Popping
Johan W. Groothoff
Jac J.L. van der Klink

Journal of Occupational Rehabiliation. In Press DOI 10.1007/s10926-009-9203-2

ABSTRACT

Objective

The aim of this study is to investigate the reliability and validity of the Disability Assessment Structured Interview (DASI). The DASI is a semi-structured interview for assessing long-term functional limitations concerning the work disability assessment of claimants.

Methods

A randomized controlled trial was conducted. Patients applying for a work-disability pension after 21 months of sick leave were independently interviewed and examined either by two physicians who had completed a DASI training period (n=32) or by two physicians from a control group (n=30) without any DASI training. Agreement percentages within both groups of physicians, eligibility for a disability benefit, and differences between the groups in terms of the scores given on the work-limitation items from the Functional Ability List (FAL) were measured to investigate reliability and concurrent validity.

To determine the content validity, the insurance physicians who completed DASI training (n=8) were asked to fill out a questionnaire concerning their opinion of the DASI. Additionally, patients filled out a questionnaire to measure their satisfaction as to the behavioral aspects of the physicians.

Results

The groups showed no important differences in agreement percentages (mean percentage about 80%) and eligibility for a disability benefit. In 9 out of 21 items the physicians of the control group indicated fewer work limitations compared to physicians using the DASI. All physicians agreed on the fact that the DASI was an acceptable tool in daily practice, one that provided a realistic picture of the patient and provided sufficient information to assess functional limitations. In addition, between the two groups, no differences were found as to the satisfaction of patients concerning the behavioral aspects of the physicians.

Conclusion

The DASI is a tool with a reasonable to good inter-rater reliability and content validity, and it appears to be acceptable to both patients and physicians. It did not improve inter-observer agreement beyond that of usual interview procedures used in the Netherlands. The DASI would seem to be a worthwhile tool for collecting self-reported information in order to assess functional limitations in claimants.

INTRODUCTION

The determination of a work disability pension for patients with long-term medical impairments is of great social and financial importance. Part of the evaluation of a work disability pension is the assessment of the functional limitations of the patient. In the literature, several instruments and their psychometric properties for assessing functional limitations are described, for instance, self-report questionnaires and performance-based functional testing[1-4]. In most countries the actual assessment of functional limitations is carried out by a medical doctor[5;6]. The assessment by physicians can be based on written information (e.g., from the patient or treating physician) or can be conducted by an examination in person. In international literature a poor agreement among physicians on functional disability exists[7-11]. A tremendous variation in disability rating recommended by physicians given the same set of facts was found[7]. To decrease this variation the United States Social Security Administration (SSA) planned to "develop functional assessment instruments that are standardized, accurately measure an individual's functional abilities ands that are universally accepted by the public, the advocacy community, and health-care professionals"[7;12;13].

In the Netherlands, an employer has to pay wages for two years if an employee is unable to work due to physical or mental disability. After these two years, the patient can apply for a work disability pension. Specialized insurance physicians assess the patient's functional limitations in work as part of the application for a work disability pension. Their judgment is based on information from treating physicians, along with their own observations, physical examination and an interview with the patient. To a large extent, the assessment is based on the interview in which attention is given to activity limitations and participation, in addition to standard medical history-taking[14]. The assessed functional limitations are registered in a standardized list, the Functional Ability List (FAL)[15].

Although in the Netherlands three semi-structured interview models for the assessment of functional limitations are available, in daily practice insurance physicians do not use a fixed model with claimants applying for disability pension[14;16]. One of the models available is the Disability Assessment Structured Interview (DASI)[17]. This is a semi-structured interview in which the three levels of functioning – impairment, activity limitation and participation – are mapped in a structured way in accordance with the International Classification of Functioning, Disability and Health (ICF)[18]. In Box 1 the general domains covered in the DASI are described.

Two important characteristics of the DASI are its semi-structured way of interviewing the patient, and its method of inquiring about specific and detailed examples of limitations and concrete activities which the patient still undertakes.

Two important criteria for evaluating work-related assessments are the validity and reliability of the instruments used[19;20]. Validity is the extent to which an instrument measures what is intended to be measured[21;22]. Content validity is the degree to which the test items represent the performance domain the test is intended to measure, and it is usually determined by a panel of experts examining the relationship between the test objectives and the test items, or by detailed knowledge of the

Box 1. Domains covered in the DASI interview

Introduction
- putting the patient at ease
- short explanation of the procedure
- summarize the known data

Work
- type and duration of work
- content of the occupation
- perception of work by the patient

Impairments
- medical history and nature of current complaints
- other diseases and general anamnesis
- course of the disease
- cause of the disease (the treating physician's diagnosis; private/work stress; personality)
- treatment and medication use (now and in the past)
- if necessary: more information from treating physician, hetero-anamnesis, expert consultation

Activity limitations
- limitations experienced in daily life and at work, for example, lifting, walking and bending.
- concrete and detailed examples of the limitations experienced.

Participation
- activities of daily living (ADL), description of a usual day, hobbies, sports, housekeeping, social contacts and work. The patient is asked which activities are actually performed and for how long.

Patient's opinion
- Patient's view of the work limitations
- Patient's response to the provisional opinion of the physician

Physician's opinion
- the physician communicates his or her judgment
- explanation of the rest of the procedure
- opportunity for the patient to respond

normal practices used. Concurrent validity examines the correlation between a new measure and an accepted measure given to the same subjects[22;23]. Reliability involves the extent to which a test or measurement is consistent and free from error[24].

In spite of the fact that assessments of functional limitations in the Netherlands are mainly based on an interview, almost no information is available on the reliability and validity of the interview as an instrument to assess functional limitations. In studies where the assessing physicians interview the patients themselves, a low inter-rater reliability was found[25;26]. In studies where physicians based their assessments on written reports or on video recordings of DASI interviews, reasonable to good inter-rater reliability was found[27;28].

Given the immense consequences of functional assessments, it is of importance to examine the psychometric properties of such an instrument. In order to fill this gap, the aim of the present study is to evaluate in a real-life situation:

- the inter-rater reliability between physicians with and without DASI training.
- the content and concurrent validity of the DASI.
- the patient's opinion of those physicians who used and those who did not use the DASI.

The DASI method was chosen over other methods because it is a well-described method which is based on the ICF and it is the only method that has had some study done on its psychometric qualities.

METHODS

Physicians

At four out of a total of 17 branches of the Dutch Social Security Office, four insurance physicians were invited to participate in the study, resulting in 16 physicians voluntarily cooperating in this study. In each of the four locations, two insurance physicians were randomly assigned to the intervention group and two were assigned to the control group. No significant difference in the average length of time spent in professional practice between the physicians in the intervention group (15.5 years, range 7-28 years) and the control group (14.6 years, range 9-21 years) was present.

Training

The intervention group was given a three-day DASI training session over a three-week period. The first week consisted of two days of instruction and practice. After demonstration of an item from the DASI by an instructor and an actress, the eight physicians practiced the items of the DASI in groups of three physicians. The role of the patient, physician and observer alternated. The next week the physicians practiced the method on their regular patients and made a video recording of the DASI interview. On the third day of the training session, in the third week, their video recordings were analyzed and assessed.

The control group did not receive any training and examined patients as usual.

Patients

A total of 443 patients who applied for social disability benefit after 21 months of sick leave were asked to cooperate, of them 236 agreed (53%). Only patients with at least lower back or lower extremity problems were selected in order to obtain a homogeneous group with sufficient filling of items of the FAL (see Instruments). Of the patients who agreed to cooperate, 26% were included (n=62), 36% were diagnosed as mental complaints (n=85), and 38% had another diagnosis such as neck and upper extremity complaints, heart and lung diseases or cancer (n=89).

Instruments

The Functional Ability List (FAL)[15] is an instrument to record functional limitations and is used in social security assessments in the Netherlands. All Dutch insurance physicians are trained and experienced in using the FAL. The FAL contains six domains in which 70 mental and physical items are addressed, and for each item the seriousness can be indicated. One example is the item "lifting or carrying", where the insurance physician has to choose from four gradations:

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 (one liter of milk)

The content validity of the DASI was assessed using a self-structured questionnaire which was filled out by the physicians who had undergone DASI training. The questionnaire contained eight questions with fixed response alternatives on a five-point ordinal rating scale. In addition, it was also possible for the physicians to make additional comments about the DASI. The questionnaire contained questions about whether the instrument was adequate for the intended purpose, whether anything essential was missing or whether any part of the instrument was irrelevant (Table 2).

In addition, the patients filled out a questionnaire that is routinely used by the Dutch Social Security Office to measure patient satisfaction with the behavioral aspects of physicians[29]. Lastly, the patients gave an indication of the duration of the interview.

Procedure

Patients (n=62) were interviewed and examined independently by two physicians from the same group (intervention group or control group) on the same day, between June and November 2008. The patients were randomly assigned to either the intervention or the control group so as to be able to compare similar groups. The physicians recorded their assessment of those work limitations to be found in the physical items of the Functional Ability List (FAL), and provided a detailed report containing information on the interview, including their judgment and the reasons for their judgment. Furthermore, we examined whether the patients did end up qualifying for a disability benefit.

After using the DASI in daily practice, the physicians were asked to give their opinions of the DASI by filling out the questionnaire. After the interview and examination, the patients were asked their opinions as to how satisfied they were with the behavioral aspects of the physicians, also by filling out a questionnaire.

Analysis

The "linear-weighted observed percentage agreement" on the FAL items was taken as a measurement of inter-rater reliability within each of the two groups of insurance physicians[30;31]. Due to the fact that the marginal distribution of the variables was skewed, the computation of an agreement index based on Cohen's kappa could not be used. One requirement for the use of this index is that the marginals should have more or less the same frequency. If not, this will result in an overestimation of the expected agreement[32]. The statistical software package AGREE 7.3[33] was used for the calculation of these values. In general, a percentage agreement of 60 to 80% is considered reasonable to good; more than 80% is considered excellent[34].

The concurrent validity was examined by comparing the mean scores on the FAL items of the intervention and the control groups. The Mann-Whitney test, a non-parametric test that is used to compare two independent groups, was used for the between-group differences in the mean scores on the FAL items.

RESULTS

A total of 62 patients were assessed by two physicians, 32 in the intervention and 30 in the control group. There were no significant differences between the groups in terms of age, gender, terms of employment and diagnosis. The mean age of the patients in the intervention group was 49.8 years (range 30–64 years), and in the control group, 46.3 years (range 35–63 years). In the intervention group 47% of the patients were female, and in the control group, 37%. Before registering sick, the patients in the intervention group worked for an average of 31.6 hours a week (range 8–40 h), and in the control group, 33.0 hours a week (range 13–40 h). In the intervention group, nine patients had lower-extremity problems (e.g., fractured ankle, gonarthrosis or peripheral arterial disease), 15 had lower-back problems (e.g., lumbar spinal stenosis, chronic non-specific lower back pain or herniated disc) and eight patients presented more general complaints (e.g., rheumatoid arthritis, fibromyalgia or somatoform disorder). In the control group, eight patients had lower-extremity problems, 14 had lower-back problems and eight had general complaints.

Table 1 presents the "linear weighted percentage agreement" between the physicians and the "mean scores" on the items of the Functional Ability List in the control and intervention groups.

Physicians from the intervention group showed a mean percentage agreement of 80.6% (range 59–100%), and the control group, 83.6% (range 67–97%). Except for the item "frequent heavy lifting," there were no differences in agreement percentages between the intervention and control groups. In 19 out of the 21 items on the FAL the physicians of the intervention group indicated more serious functional limitation scores in their assessments compared to the control group. For nine of these items, there were significant differences ($p < 0.05$). Concerning the daily number of hours a patient could function, the physicians in the intervention group gave limitations in 31% of the patients; in

Table 1. "Linear weighted percentage agreement" between the physicians (columns 1–2) and "Mean scores" (columns 3–4) on items of the Functional Ability List in the intervention (n=32) and control (n=30) groups

Items	Agreement (%)		Mean ± SD (min-max)	
	Intervention	Controls	Intervention	Controls
Body movement scale				
Reaching	97	97	0.02±0.13 (0-1)	0.02±0.13 (0-1)
Frequent reaching	81	88	0.56±0.73 (0-2)	0.55±0.79 (0-3)
Bending (degrees)	78	85	0.75±0.74 (0-2)*	0.48±0.68 (0-2)
Frequent bending	83	86	1.50±0.85 (0-3)	1.35±1.01 (0-3)
Rotation¶	78	87	0.23±0.43 (0-1)*	0.07±0.25 (0-1)
Push or pull	75	72	1.09±0.68 (0-2)*	0.62±0.61 (0-2)
Lifting or carrying	85	80	1.62±0.75 (0-3)*	1.07±0.69 (0-2)
Frequent light lifting	78	84	0.73±0.93 (0-3)	0.60±0.81 (0-3)
Frequent heavy lifting¶	100*	83	1.00±0.00 (1-1)*	0.88±0.32 (0-1)
Walking	91	86	1.30±0.79 (0-3)	1.08±0.85 (0-3)
Sustained walking	90	87	1.53±0.69 (0-3)*	1.20±0.69 (0-2)
Climbing stairs	79	78	1.22±0.93 (0-3)*	0.78±0.69 (0-2)
Climbing	70	87	1.08±0.78 (0-2)*	0.80±0.55 (0-2)
Kneeling¶	72	67	0.16±0.36 (0-1)	0.10±0.30 (0-1)
Body posture scale				
Sitting	86	88	0.98±0.85 (0-3) *	0.55±0.65 (0-2)
Prolonged sitting	79	90	0.78±0.70 (0-3)	0.57±0.56 (0-2)
Standing	80	82	1.55±0.81 (0-3)	1.40±0.80 (0-3)
Prolonged standing	88	89	1.59±0.71 (0-3)	1.47±0.72 (0-3)
Prolonged kneeling¶	72	87	0.48±0.50 (0-1)	0.63±0.49 (0-1)
Prolonged bending¶	59	80	0.58±0.50 (0-1)	0.43±0.50 (0-1)
Working above shoulder¶	72	73	0.27±0.48 (0-1)	0.36±0.45 (0-1)
Mean	**80.6**	**83.6**		

¶ = dichotomous data, other items are ordinal; * $p < 0.05$

40% of these patients the physicians were in agreement on this. In the control group, the physicians indicated a limitation in hours of daily functioning of 23% in their patients; in 29% of these patients the physicians were in agreement on this.

In the intervention group, 18 out of 32 patients (56%) qualified for a work disability benefit, while in the control group 13 out of 30 patients (43%) did; this did not represent a significant difference ($p = 0.31$).

Table 2 presents the opinion of the eight physicians of the intervention group concerning the DASI. All physicians were in agreement that the DASI was an acceptable tool in daily practice, one which gives an objective view of the patient and enough information to assess functional ability.

Table 2. The physicians' opinions (n=8) on the DASI (intervention group) in percentages

The DASI …	disagree %	neutral %	agree %
… provides an objective view of the patient			100
… attends to all relevant aspects		12.5	87.5
… attends to irrelevant aspects	87.5	12.5	
… is acceptable in daily practice			100
… provides enough information to assess functional ability			100
… differs from my own interview	25	25	50
… has added value		12.5	87.5
… allows a better founding for filling out the FAL[1]	12.5		87.5

[1] Functional Ability List

As an added value of the DASI, the physicians mentioned in particular the structuring of the interview and collecting detailed information on the functioning of the patient. One physician mentioned that the DASI mainly collected information from the patient, but that the assessment of this information into functional abilities was not addressed.

In their reports the physicians of the intervention group mentioned an average of 6.7 functional limitations as experienced by the patient (range 4–10). In the control group, an average of 4.4 functional limitations were mentioned (range 0–7) ($p < 0.05$). In the case of functional limitations, 71% of the intervention group indicated the intensity of the limitations experienced, for instance, by giving an example of the limitation in daily life. In the control group, this was 40%.

The patients' satisfaction report score for physicians of both the intervention and control groups in their interviews was 7.7 on a scale from 1 to 10. Moreover, no differences between the two groups were found in terms of answers to the questions concerning behavioral aspects of the physicians (listening, empathy, meticulousness and professionalism).

According to the patients, the duration of the interview and the physical examination was on average 45–60 minutes in the intervention as well as in the control group (range < 30 min – > 60 min).

DISCUSSION

Although accurate determination of work disability status is crucial for the health and well-being of patients and their families, the reliability and validity of bureaucratic approaches is poor and cumbersome and extraordinarily expensive. This study of the DASI demonstrates that a semi-structured interview might hold great promise as an inexpensive solution to this problem. We studied inter-rater reliability, and both content and concurrent validity, along with the patient's opinion of the DASI.

Reliability

Up till now, no real life studies of inter-observer agreement among physicians in assessing functional limitations had been conducted. We hypothesized that agreement between physicians in the control group would be low and, in the DASI group, that it would be acceptable. In the end, we found an overall inter-rater reliability for the items of the FAL in the intervention group that was reasonable to good, and for some dimensions even excellent. Contrary to our expectations, the agreement in the existing practice was satisfactory too, and DASI training did not improve agreement between physicians. One explanation for this may be the fact that we used patients with relatively straightforward lower back or lower extremity problems. Possibly those patients with more complicated problems and those with mental problems might produce less satisfactory results. Because agreement between physicians in international literature is found to be very poor[7-11], another explanation may be that the satisfactory agreement in existing practice is specific for the Dutch context. In the Netherlands specially trained insurance physicians assess the functional disabilities in patients. These physicians all had an interview-training, in which they were taught to ask for activity limitations and participation in addition to standard medical history-taking. This education is not always common in other countries, and may be the explanation of the relatively good agreement between the physicians.

We found a low inter-rater agreement concerning the daily number of hours a patient could function. The daily number of hours a patient can function according to the physician often has very important consequences for a work disability benefit. Therefore, the low inter-rater reliability found is undesirable. Insurance physicians in the Netherlands have a guideline for "reduced working hours"[35] at their disposal, but unfortunately this guideline cannot prevent the differences in outcome between the physicians. The satisfactory inter-rater reliability on the items of the FAL and the low inter-rater agreement concerning the daily number of hours a patient could function which were found in this study are comparable to Dutch studies conducted in a more controlled environment where physicians did not see patients face to face, but made an assessment based on video recordings or written reports of DASI patient interviews[27;28].

Validity

Preferably, validity is assessed by comparing the measurement studied to a gold standard. For assessing functional limitations, however, no gold standard is available. Different methods for assessment, for instance, self-assessment questionnaires, clinical examination and performance tests, lead to different outcomes[36]. From the reports made by the physicians in this study, it appeared that the same information could lead to different outcomes. One example was the assessment of a 56-year-old patient with depression and lower-back problems as a result of a somatoform disorder. One physician assessed no functional limitations when considering the diagnosis and an absence of objective functional defects. The other physician assessed the same patient and concluded the patient was limited in lifting ability (10 kg maximum), sitting (1 h maximum) and walking (half an hour maximum) because the patient made a genuine impression, and offered a plausible and consistent story. The question might be raised as to whether consistency in a patient's behavior together with the functional limitations experienced should in fact be leading factors in the assessment, this despite the fact that there might be no actual objective medical findings present. In this light, part of the assessment of functional limitations would seem to lie in the realm of a social rather than a medical concept.

This study showed a satisfactory content validity for the DASI. Without a single exception, the physicians agreed on the fact that the DASI was an acceptable tool in daily practice and one which gave an objective view of the patient and enough information in order to assess functional abilities. Seven out of eight physicians found the DASI to be an even better basis for the assessment of functional limitations than the interview they usually applied.

For assessing concurrent validity, the outcomes of two measurements administered to the same patients were compared. In this study, we compared the outcomes of the intervention group and the control group in different groups of patients. Because patients were randomly allocated to the intervention group and the control group, however, the groups were comparable. This is supported by the fact that there were no significant differences for both groups in terms of age, gender, terms of employment and diagnosis. We found that in almost half of the items of the FAL, the physicians using the DASI gave substantially more severe functional limitations in their assessments than did the control group. But this did not lead to an increased number of patients who qualified for a disability benefit. One explanation for the more severe functional limitations may be that the DASI focuses more attention on problems concerning activities and functional limitations as compared to "care as usual." That this did not lead to more disability benefits can be explained by the Dutch system for determining the benefit. An occupational expert investigates what jobs the patient is theoretically still able to perform in light of the limitations. The earning capacity will determine the disability benefit. Apparently the more severe functional limitations did not lead to a greater loss in earning capacity. The literature describes the fact that insurance physicians show limited attention to the detailed information regarding the functional limitations the patients experience[14]. This is in line with the findings in this study where the physicians in the intervention group reported

significantly more severe functional limitations.

Even though no differences between the intervention and the control groups were found in disability benefit outcome, the difference in outcome for the functional limitations was important because functional limitations are needed for reintegration into appropriate work. Because of the lack of a gold standard, it is unknown whether the more severe ratings in the DASI group are more valid than those of the "usual care" physicians. The physicians in the intervention group found the DASI to be a better basis for the assessment than the interview they usually applied. Therefore, we think the DASI contributes to a more thorough assessment of the functional limitations.

Patients' opinions

The patients' mean report score for satisfaction with the DASI was 7.7 on a scale from 1 to 10. The same score was found for the interviews of the physicians in the control group. Apparently the DASI did not improve or worsen patient satisfaction.

Study limitations

Some limitations of this study should be noted. The physicians knew they were being monitored; this might have influenced their assessments. However, it was practically impossible to conduct a study which would not have had this disadvantage. Furthermore, the assessments were aimed at physically based functional limitations; mentally based functional limitations might well present a rather different outcome. Finally, although the physicians in the intervention group received DASI training, it is possible they did not implement this in daily practice. We studied the reports of the assessments to check whether the physicians who received DASI training actually performed the interview as it was taught. One important characteristic of the DASI is the presence of concrete and detailed information on functional limitations as experienced by the patient. The reports on those physicians who received the training contained more functional limitations and more detailed information on this point, indicating that the intervention group actually performed what they had been trained to do.

The DASI in daily practice

Several tools are used to assess the functional abilities of people with medical impairments, but no single currently existing test provides a valid measurement of functional limitations[12;37]. Functional capacity tests and questionnaires alone cannot properly assess functional limitations without an appraisal of the outcome of these tests. A combination of specialized physicians and instruments such as functional capacity tests and questionnaires looks the most promising. It might be useful to provide self-report questionnaires about function to the patient before the DASI in order to increase the efficiency and specificity of the interview. Then, clinical examination and a semi-structured interview, like the DASI, could be conducted by a physician or, in part, even by a trained nurse[28]. Based on this information, individually selected functional capacity tests could be conducted to

confirm or disconfirm the initial results of the interview. Guidelines and protocols might narrow down further the differences in assessment among physicians[25].

Future research

Further research into the value of guidelines and protocols – especially where the assessment of limitations as to the number of hours a patient can function daily is a factor – as well as additional studies concerning the use of the DASI in mental-function limitations may be useful. Concurrent validity can be assessed by comparing outcomes of the DASI with self-report questionnaires and functional capacity tests. For research into the validity of instruments to assess functional limitations, a gold standard is needed. A gold standard might be approached by looking for consensus among a number of physicians after medical examination, an interview protocol, questionnaires and performance tests.

CONCLUSION

In conclusion, we would state that the DASI is a tool with a reasonable to good inter-rater reliability and content validity, and that it appears to be acceptable to both patients and physicians. The DASI did not improve inter-observer agreement beyond that of usual interview procedures used in the Netherlands. The DASI would seem to be a worthwhile tool for collecting self-reported information in order to assess functional limitations in claimants. Because the physicians who used the DASI assessed more functional limitations as compared to usual practice, further research into the interpretation of the self-reported information is needed.

Reference List

1. Bot SDM, Terwee CB, van der Windt DAWM, Bouter LM, Dekker J, de Vet HCW. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. Annals Of The Rheumatic Diseases 2004;63:335-341.
2. 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;3:130-140.
3. Wind H, Gouttebarge V, Kuijer PP, Frings-Dresen MH. Assessment of functional capacity of the musculoskeletal system in the context of work, daily living, and sport: a systematic review. J Occup Rehabil 2005;15:253-272.
4. Matheson L, Mayer J, Mooney V, Sarkin A, Dreisinger T, Verna J et al. A method to provide a more efficient and reliable measure of self-report physical work capacity for patients with spinal pain. J Occup Rehabil 2008;18:46-57.
5. De Boer WEL, Besseling JJM, Willems JHBM. Organisation of disability evaluation in 15 countries. Pratiques et Organisations de Soins 2007;38:205-217.
6. Wahlstrom R, Alexanderson K. Chapter 11. Physicians' sick-listing practices. Scandinavian Journal of Public Health 2004;32:222-255.
7. Clark W, Haldeman S. The development of guideline factors for the evaluation of disability in neck and back injuries. Spine 1993;18:1736-1745.
8. Söderberg E, Alexanderson K. Sickness certification practices of physicians: a review of the literature. Scandinavian Journal of Public Health 2003;31:460-474.
9. Lax MB, Manetti FA, Klein RA. Medical evaluation of work-related illness: evaluations by a treating occupational medicine specialist and by independent medical examiners compared. International Journal of Occupational & Environmental Health 2004;10:1-12.
10. Wolfson AM, Doctor JN, Burns SP. Clinician judgments of functional outcomes: how bias and perceived accuracy affect rating. Archives of Physical Medicine & Rehabilitation 2000;81:1567-1574.
11. Rainville J, Pransky G, Indahl A, Mayer EK. The physician as disability advisor for patients with musculoskeletal complaints. Spine 2005;30:2579-2584.
12. Matheson LN, Kane M, Rodbard D. Development of new methods to determine work disability in the United States. J Occup Rehabil 2001;11:143-154.
13. Social Security Administration. Plan for a new disability claim process. Washington, DC: Social Security Administration, 1994.
14. 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, the Netherlands: CTSV, 2000.
15. Functional Ability List. ClaimBeoordelings- en BorgingsSysteem (CBBS). Amsterdam: Lisv, 2002.
16. De Boer WEL, Wijers JHL, Spanjer J, Beijl I van der, Zuidam W, Venema A. Gespreksmodellen in de verzekeringsgeneeskunde. [Interview models in insurance medicine]. Hoofddorp, the Netherlands: TNO, 2004.
17. Spanjer J. Disability Assessment Structured Interview (DASI), description of an assessment method. Zuidlaren, the Netherlands, 2007. A copy can be acquired from the author.
18. International classification of functioning, disability and health. Geneva: World Health Organisation, 2001.
19. Innes E, Straker L. Attributes of excellence in work-related assessments. Work 2003;20:63-76.
20. Matheson LN, Gaudino EA, Mael F, Hesse BW. Improving the validity of the impairment evaluation process: A proposed theoretical framework. J Occup Rehabil 2000;10:311-320.

21. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. 2 ed. Upper Saddle River, New Jersey, Prentice Hall, Inc, 2000.
22. Innes E, Straker L. Validity of work-related assessments. Work 1999;13:125-152.
23. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. 2nd edn. Oxford: Oxford University Press, 1995.
24. Innes E, Straker L. Reliability of work-related assessments. Work 1999;13:107-124.
25. Spanjer J. De reproduceerbaarheid van WAO beoordelingen, een literatuuronderzoek. [The reproducibility in disability assessment: a literature search]. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2001;9:195-198.
26. Brouwer S, Dijkstra PU, Gerrits EHJ, Schellekens JMH, Groothoff JW, Geertzen JHB et al. Intra- en inter-beoordelaarsbetrouwbaarheid 'FIS-belastbaarheidspatroon'en Functionele Mogelijkheden Lijst". [Intra- and inter-rater reliability 'Functional Information System' and 'Functional Ability List']. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2002;11:360-367.
27. Spanjer J. De betrouwbaarheid van WAO beoordelingen. [The inter- and intra- rater reliability of disability pension assessments]. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2001;9:234-241.
28. Spanjer J, Krol B, Brouwer S, Groothoff JW. Inter-rater reliability in disability assessment based on a semi-structured interview report. Disability & Rehabilitation 2008;30:1885-1890.
29. Burg CLvan der. Clientmonitor. Leiden, the Netherlands: AStri, 2007.
30. Schouten HJA. Measuring pairwise agreement among many observers. Biometrical Journal 1980;22:497-504.
31. Schouten HJA. Measuring pairwise agreement among many observers. II. Some improvements and additions. Biometrical Journal 1982;24:431-435.
32. Popping R. On agreement indices for nominal data. In: Saris W.E., Gallhofer I.N. Sociometric Research. Volume 1. Data collection and scaling. New York: St. Martin's Press, 1988, pp. 90-105.
33. Popping R. AGREE, A package for computing nominal scale agreement. Comput Stat Data Anal 1984;2:182-185.
34. Landis JR, Koch GG. The measurements of observer agreement for categorical data. Biometrics 1977;33:159-174.
35. Standaard verminderde arbeidsduur. [Guideline reduced working hours]. Amsterdam: Lisv, 2007.
36. 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. Disability & Rehabilitation 2005;27:999-1005.
37. Gaudino EA, Matheson LN, Mael FA. Development of the Functional Assessment Taxonomy. J Occup Rehabil 2001;11:155-175.

CHAPTER 8

General discussion

INTRODUCTION

In the Netherlands, for two years employees are still paid their salary by their employer when disabled for work. After two years, the employee can then apply for workers' compensation benefit from the Social Security Office. In awarding a workers' compensation benefit, the insurance physician plays an important role in the work-disability assessment. The insurance physician assesses mental and physical limitations and the abilities of the patient and records them in a standardized list – the Functional Ability List (FAL) (Appendix 1)[1]. To assess these functional limitations, insurance physicians use different sources of information including information from the patients themselves, from the patients' treating and occupational physicians, and from observation and physical examination. Most of the information is gathered during a face-to-face interview between the insurance physician and the sick-listed employee[2]. Despite the major consequences of work-disability assessments, no studies can be found on the reliability and validity of these assessments and the instruments used[3]. In the first part of this thesis, a model will be presented in which an overview is shown of the different steps insurance physicians have to take in assessing the functional limitations, along with the possible instruments they can use. In the literature, several instruments are described for assessing functional limitations. Most of these instruments are used in a rehabilitation setting to assess limitations of patients in their daily lives. These instruments cannot be used in disability assessments as is, because a rehabilitation setting differs from a claim setting, and limitations in daily life are quite different from limitations in work. Therefore, we have conducted a systematic literature search to find instruments to assess functional work limitations in claimants.
In the Netherlands, the interview plays an important role in disability assessment as a method of gathering information about clients' functional limitations. Therefore, in the second part of this thesis we have studied the reliability and validity of, and clients' and insurance physicians' satisfaction with one specific interview model: the Disability Assessment Structured Interview (DASI)[4]. The DASI is one of three interview models that are described and taught to insurance physicians as part of their education. In daily practice, insurance physicians often do not use one single model, but will use parts of the three different models[5]. The DASI is a semi-structured interview with specific attention paid to concrete and detailed examples of functional limitations and activities (Appendices 2 and 3).

In this chapter, we will summarize the main findings of this thesis, discuss these findings and their methodological issues, comment upon the practical implications, and then make recommendations for further research, practice and policy.

MAIN FINDINGS

The main findings can be summarized by answering the research questions formulated in the first chapter:

1 *What are the possible sources of variation in work-disability assessment?*
Variation in assessing functional limitations may be come about at three different levels: data collecting, interpretation, and documentation of these data. In all of three of these steps the qualities of the assessor play a major role (Chapter 2).

2 *Which instruments are described that measure or assess functional limitations in claimants, and what are their psychometric qualities?*
We did not find any instruments with satisfactory psychometric qualities for assessing functional limitations in work disability claimants (Chapter 3).

3 *What effect does detailed information on functioning in addition to medical history-taking have upon the functional limitations assessed and on inter-rater reliability?*
Detailed information on functioning provided by the patient has no important influence on the inter-rater reliability of the functional limitations assessed.
However, there are significantly higher scores on assessed functional limitation items as compared to medical history-taking alone (Chapter 4).

4 *In their own opinion, are physicians able to assess functional limitations based on a written DASI report?*
Half the physicians thought that a reliable assessment based on the written information was possible, 25% found the opposite, and 25% was indecisive (Chapter 5).

5 *What are the characteristics of the DASI interview in daily practice?*
In our study in 2000, we measured a mean interview duration of 33 minutes (range 19-77 min). In the total interview, 9% was spend on the introduction, 9% on work perception, 30% on impairments, 29 % on functional limitations and activities, 8% on the client's opinion, 13% on the physicians the decision, and 4% on a final reaction of the patient. (Chapter 6).
In our study in 2008, patients indicated that the duration of the DASI interviews and the physical examination was on average 45-60 minutes (range < 30 min – > 60 min) (Chapter 7).

6 *What is the patient satisfaction for physicians who conduct a DASI interview?*
The patients' satisfaction report score for the physicians was 7.7 on a scale from 1 to 10. No

difference between physicians who conducted DASI interviews and a control group was found (Chapter 7).

7 *What comments on the DASI do insurance physicians have?*
The quality of written DASI interview reports was evaluated as reasonable to good, but sometimes more information (physical and mental examination, information from treating physician) was needed for an assessment. Video-taped DASI interviews were described as structured and efficient. However, more medical (medical history, complaints, and therapy) and social (family) information was found to be needed. Physicians who conducted DASI interviews themselves found the DASI to be an acceptable tool in daily practice (Chapter 5-7).

8 *What is the opinion of insurance physicians when using the DASI?*
Physicians using the DASI in daily practice found that the DASI provided a realistic picture of the patient and provided sufficient information to assess functional ability (Chapter 7).

9 *What is the intra- and inter-rater reliability of functional limitations assessments using the DASI?*
The intra- and inter-rater reliability in DASI assessments in general was moderate to good, but poor on the item "hours a patient can work daily." Inter-rater reliability was no better than for a control group who did not have any special training (Chapter 5-7).

10 *What is the content and concurrent validity of functional limitations assessments using the DASI?*
Content validity was satisfactory: All of the physicians found the DASI provided sufficient information to assess functional limitations and that it attended to all relevant aspects. Physicians using the DASI indicated more serious functional limitations compared to usual practice. However, this did not lead to an increased number of patients who qualified for a disability benefit (Chapter 7).

DISCUSSION

Despite the fact that the assessment of functional work limitations worldwide is an important issue[6], we found that almost no validated and reliable instruments for this assessment are described. Given the importance of these assessments individually and socially, this represents an important gap between disability research and daily practice. Several instruments can be used to assess functional limitations, including performance tests, self-assessment questionnaires and clinical examination, but the psychometric properties of these instruments in workers' compensation claimants have scarcely been studied before at all. It is remarkable that instruments frequently used world-wide,

such as self-assessment questionnaires and performance tests, barely play any role in disability assessment in the Netherlands. These instruments might possibly prove to have additional value in the Netherlands, in addition to the interview. Questionnaires might, for instance, be useful as basic information prior to the interview and examination by the insurance physician. After the physician's examination, questionnaires could be used for specific indications, for instance, a questionnaire to measure depression or mental limitations. Performance tests could also be used for specific indications after the examination in order to obtain supplemental information about functioning[7]. Because the anamnesis of the patients plays an important role in the assessment in the Netherlands, we studied the psychometric properties of one interview model, the Disability Assessment Structured Interview (DASI).

DASI – general issues

The DASI is an interview in which the three levels of functioning as described in the International Classification of Functioning, Disability and Health (ICF)[8] – impairment, activity limitation, and participation – are mapped out in a structured way. We found that a considerable part of the DASI interview was spent on activity limitations and participation (about 30% of the total interview). We also found that this was distinctly more than in usual practice. Insurance physicians are required to assess functional limitations; therefore, it would seem logical that a specific anamnesis into functional limitations should play a prominent role. This was also the opinion of insurance physicians who had undergone DASI training; they indicated that this specific anamnesis into functional limitations had added value as compared to their own anamnesis. As part of patient empowerment, this specific anamnesis would also seem to be useful, because this way the patient would have the opportunity to indicate which limitations are experienced and to what extent. The patient satisfaction, for that matter, was no different than in usual care. This was probably because this usually depends more on the attitude of the assessing physician than on the content of the questions. This is in line with a study in which the satisfaction of the patient depended mainly on the explanation and listening capabilities of the insurance physician[9].

We found differences in the mean duration of the DASI interview in a study done in 2001 (33 minutes) and in 2008 (45-60 minutes). Several explanations could be found for this difference. First, between 2001 and 2008 the Disability Act changed (WAO became WIA in 2006) and patients were assessed after two years of disability instead of after one year. In addition, more pressure was placed on employers to reintegrate employees into work. As a consequence, more complex patients had to be assessed, who took more time to interview. Moreover, the duration of the DASI interviews in 2001 had been measured with a stopwatch, while in 2008 patients estimated the duration of interview and additional examination (about 10 minutes). Generally speaking, a mean duration of a present-day DASI interview without physical examination can be assumed to be 45 minutes.

To study whether physicians could assess functional limitations based on written information alone, we asked physicians if they thought they could make a reliable assessment based on a

written DASI report. Only half of the insurance physicians thought this was possible. The physicians mostly indicated that additional information concerning observation and physical examination was needed. If the insurance physicians saw the patient themselves in the consulting room or on a video recording, they indicated that a proper assessment was possible. Apparently, a face-to-face contact is useful because the physician not only bases the assessment on the content of the interview, but also on the observation of the patient.

DASI – reliability

In this thesis the reliability of disability assessments using the DASI was measured by measuring the agreement among physicians with respect to functional limitations on standardized lists. In the international literature, there is poor agreement among physicians as to functional disability, along with a tremendous variation in the disability rating recommended by physicians given the same set of facts was found[10-14]. In the Netherlands, an important variation among physicians in disability assessment was presumed[3]. We found that the inter-rater reliability of most items of the FIS and FAL was moderate to good regardless of the interview method that was used. The relatively satisfactory agreement among physicians on the assessment of functional limitations in the Netherlands can be explained by the Dutch system which includes the use of specialized insurance physicians, guidelines and protocols, mutual agreement meetings, and perhaps the use of the interview, as the most important instruments in assessing limitations.

Contrary to what might be expected, however, inter-rater reliability did not improve if physicians had the same (DASI) interview training. In the interview, the physician first collects information on the patient, and then this collected information is interpreted. We found that the same information could lead to different assessments because of differences in interpretation. This was clearly present when assessing "hour limitations," but also in diseases without objective medical findings. An example is a patient with psychosomatic complaints in which one physician assessed no limitations because there was no physical reason, while another found there were limitations because the patient made a genuine impression, and offered a plausible and consistent story. Therefore, inter-rater reliability seems to benefit more from a univocal interpretation of information, for instance, by administering proper guidelines and protocols, rather than by a univocally administered collecting of information in the interview.

The low agreement among physicians on the item "hours a patient can work daily" that was found in this thesis is a serious problem, because it is one of the most important items in work-disability assessment. If there are limitations on this item, very often this will lead to (partial) work disability benefit. On this item a special "reduced hours" guideline[15] exists, which states that the physicians have to assess the amount of hours a patient can theoretically function a day, and that other duties patients have in their private lives (e.g., parenting and household) cannot be taken into account. However, it is difficult to distinguish this theoretical concept from the amount of hours a patient can work in practical terms, and in combination with private duties. Given the differences found on this

item, a further consideration of this item would seem to be appropriate, for instance, by creating a new guideline or doing a better job of implementing the present one.

DASI – validity

In this thesis, the validity of disability assessments using the DASI was measured in two ways. First, by asking the opinion of insurance physicians concerning the DASI (content validity) and, second, by comparing the assessment outcome of physicians who had DASI training with the outcome from usual practice (concurrent validity). Content validity was satisfactory: Without a single exception, the physicians agreed on the fact that the DASI was an acceptable tool in daily practice, and one which gave an objective view of the patient and enough information in order to assess functional abilities. In comparing usual practice to assessments with the use of the DASI, however, a significant difference was found. In assessments where the DASI was used, the physicians indicated substantially more severe functional limitations. This can be explained by the fact that in the DASI interview physicians are specifically trained to inquire after concrete and detailed examples of problems in functioning. Compared to usual practice, more information on functioning is gathered in the DASI interview. This was found in the video study were the physicians mentioned that in their own anamnesis they asked for fewer details on functioning, but also, in the study where physicians had undergone DASI training, they noted more functional limitations in their reports than in usual practice. We also found that physicians using detailed information on functioning indicated more serious functional limitations than did physicians who received medical information alone. Apparently, if physicians base their assessments on medical facts rather than on daily functioning, fewer serious functional limitations are assessed.

If this is true, patients are possibly performing on a lower level than is medically necessary, and acknowledging this behavior can cause medicalization by granting the patient a disability benefit. Since we found no increased number of patients who qualified for a disability benefit compared to usual practice, we think this problem is negligible. The advantage of complying with the limitations the patient experiences can result in a better acceptance of the assessment by the patient. A consequence of this might be fewer appeal cases and easier reintegration into appropriate work. Another possible explanation is that patients don't perform on a lower level than is medically necessary, but that those physicians who based their assessments on medical facts rather than on daily functioning underestimated the functional limitations of the patients.

To investigate the correct functional limitations, a gold standard is needed. For assessing functional limitations, however, no gold standard is available. The lack of a gold standard is caused by the fact that there is no clear relationship between disease and functional limitations[16;17]. Often patients experience functional limitations because of pain or discomfort, but no medical reasons for these functional limitations are present, because no physical damage is present. In the International Classification of Functioning, Disability and Health[8], performance (what a patient does) and capacity (what a patient can do) are distinguished. In daily practice there is often a gap between

performance and supposed capacity; mostly performance is lower than capacity. One example is patients with "non-specific chronic low-back pain" in which there are no medical reasons why the patient is limited in his actions (capacity). Despite this, a lot of patients restrict themselves when lifting, bending or sitting because they experience pain (performance). Different physicians can have different opinions about what needs to be measured: capacity, performance or a combination of the two[18]. To resolve this problem, in the Netherlands the "medical disability criterion" is legally documented[19]. According to this criterion, the patient's experience should be the starting point for the assessment. However, a loss of autonomy must be present, along with the requirement that impairments, activity limitations and handicaps should fit together consistently. Unfortunately, as a result, there still is no agreement among insurance physicians in their assessments.

METHODOLOGICAL ISSUES

Study design

Finding an adequate design for studying the reliability and validity of assessments using a semi-structured interview model is difficult for several reasons. First, although the interview makes up a very important part, it is not the interview alone which insurance physicians base their assessments upon. For instance information in the patient's file, observation of the patient, physical examination, medical knowledge, and mutual agreements among physicians also play a role. Therefore, elements other than the interview model alone will bias the outcome in terms of reliability and validity. Physicians always need other kinds of information in addition to the interview alone, because they have to check whether the information the patient provides is consistent with that other information, and they need specific medical information that the patient cannot provide. Therefore, a study design without this bias is not possible.

Studying the inter-rater reliability among physicians in patient assessments would be problematic for several reasons. Two or more physicians would have to assess the same patient. Preferably, this should happen at about the same time, because the functional limitations can change over time and then physicians would end up not assessing the same limitations. If physicians see patients right after each other, it is possible that the consultation with the first physician might affect the consultation with the second physician. For instance, if the first physician in the interview asks for detailed information on functional limitations several times, it is possible that the patient will spontaneously report this in the interview with the second physician. Furthermore, an interview and examination would be stressful for a patient, and two or even more consultations after each other might be too much for patients. As a result, finding patients who are prepared to cooperate with such a study might well be a problem.

A written report or a video recording of a patient both have the advantage that several physicians can assess the same patient at the same time, and only one contact with the patient is needed. One disadvantage is the fact that the inter-rater reliability score may be higher than in daily practice,

because there is no patient-physician interaction. One physician can ask other questions of the patient and obtain different information than another physician, which can result in a different assessment. Therefore, the design of an adequate study concerning inter-rater reliability is not simple. Our choice was three different study-designs: written reports, video recordings and real-life assessments. To study the validity of the DASI is a problem because no gold standard exists. However, there are other possibilities for assessing the validity of the DASI. For instance, the outcome of the DASI can be compared with the outcome from other ways of assessing functional limitations such as performance tests, questionnaires or other interview protocols. One problem, though, with all these kinds of studies is that what the patient shows in a laboratory situation or what the patients say then has to be relied on. An alternative would be covert observation of the patient in order to observe the patient's real activities in daily life[20]; however, ethical objections could be expected with such a kind of study. In this thesis, though, it was decided to measure the validity of disability assessments using the DASI in two ways. First by asking the opinion of insurance physicians concerning the DASI (content validity) and second by comparing the outcome on the FAL of physicians who had DASI training with that of usual practice (concurrent validity).

Outcome measurements

We used the Functional Information System (FIS), Functional Ability List (FAL) and Mental Ability List (MAL) as outcome measurements in the different studies. One problem with these instruments is that reliability, validity and other psychometric properties have not been demonstrated yet. Nevertheless, because of the lack of alternatives, we chose to use these instruments because insurance physicians are trained and experienced in using the FIS and the FAL, since they are used in all work-disability assessments in the Netherlands. The MAL is an instrument for recording mental limitations and has been developed through daily practice. Although the MAL is not used in all disability benefit assessments, as the FIS and FAL are, we chose to use this instrument because it was a simple instrument which needed no training to use, and gave a clear view of mental abilities. Face and content validity of the three instruments seem to be sufficient because the instruments are developed after study of the relevant literature, years of comments and feedback from insurance physicians, and extensive use in daily practice, along with the important demands that are relevant to daily work functioning are described. It is striking, however, that no studies into the psychometric properties of such an important instrument such as the FAL have been conducted.

Statistical analysis

To measure intra- and inter-rater reliability, we used agreement percentages on the items of the FIS, FAL and MAL. This is not an ideal measurement to use, because no correction for chance is applied. For instance, if the agreement percentage in insurance physicians on the dichotomous FAL item "heavy lifting" is 92%, this would seem to be a very good result. However, if this concerns a study on low-back pain patients, 90% of whom have trouble in heavy lifting, the agreement among the

physicians for the most part could be explained simply by chance.

Cohen's kappa[21] is a measure that compares agreement against that which might be expected by chance. The kappa ratio is a proportion that can go from -1 (perfect disagreement) through 0 (agreement expected by chance), on to +1 (perfect agreement)[22]. Unfortunately, one requirement for the use of the kappa is that the marginals should have more or less the same frequency. If not, this will result in an overestimation of the expected agreement, and the kappa's will be very low[23]. Due to the fact that the distribution of the FIS and FAL items were skewed, Cohen's kappa could not be used.

In dichotomous data, absolute agreement percentages can be calculated. Because most items of the FIS, FAL and MAL were ordinal, and involved three or more categories, we used a weighted percentage agreement. The weighted agreement gives partial credit for agreement that is close but not perfect. For instance, if on the item "walking," Physician A scores that the patient can walk for two hours, Physician B for one hour and Physician C for 15 minutes, the agreement between Physicians A and B is higher than the agreement between Physicians A and C[24-26].

Comparing results

It would be tempting to set up a table in which the results in terms of inter-rater reliability in the different studies in this thesis were collected and compared. However, the different studies cannot were not really comparable because:

- The DASI was adjusted as a result of the comments of the insurance physicians in the different studies. For instance, the item "social contacts and functioning" was introduced only in later versions of the DASI.
- In the written report and video studies, MAL and FIS items were used as outcomes, while in the "real-life" study the FAL was used. The FIS often has more grades in the different scales, and in earlier studies better agreement percentages in the FAL than in the FIS were found[27].
- In the real-life study only patients with low back or lower leg problems were used, whereas in the written report and video studies the patients had several different physical and mental diagnoses.
- The legislation changed in between the studies. In April 2002, the "Improved Gatekeeper's Act" was introduced and, in January 2006, a new Dutch Disability Act. The result was that patients were assessed after two years instead of after one year of disability, and assessments were more complicated.

PRACTICAL IMPLICATIONS

In daily practice, insurance physicians often use no single interview protocol, but instead use parts of different protocols[5]. A single interview protocol for all insurance physicians in the Social Security Office in the Netherlands would seem advisable, because it would result in a uniform disability

assessment. It may be too early to recommend the DASI as this state-of-the-art interview protocol, because further research into its validity is needed. The outcome of the DASI especially needs to be compared with some kind of "silver standard," given the lack of a gold standard. Such a silver standard might be approached by looking for consensus among a number of physicians after the medical examination, an interview protocol, and questionnaires and performance tests.
Although it would seem to be too early to use the DASI as a state-of-the-art interview protocol, there are sufficient arguments present to offer DASI training in the education of insurance physicians and as a refresher course for insurance physicians already registered. These arguments include:

- Experienced insurance physicians are satisfied with its use in daily practice. They even think it is a better basis for assessment of functional limitations than their own interview.
- It has satisfactory content validity; insurance physicians found that all relevant aspects and no irrelevant ones were attended to.
- It seems logical to ask patients in detail which functional limitations they encounter in daily life, when it is their functional limitations which need to be assessed.
- Physicians using the DASI interview indicated more serious functional limitations compared to usual practice. Though this has no effect on the eligibility for a compensation benefit, complying with the functional limitations patients experience can have consequences for their reintegration into other jobs. If functional limitations are underestimated, patients may start in a new job which is too taxing for them.
- In this thesis, acceptable reliability, validity, and patient and physician satisfaction were demonstrated. We found no alternatives that had better psychometric qualities for assessing functional limitations in workers' compensation claimants.

The Netherlands School of Public and Occupational Health (NSPOH), which educates all insurance physicians in the Netherlands, offers a short introduction to the DASI method at this point in time, one which takes a maximum of one day. Our experience tells us that this is really too short and that the three-day training as described in our study is a better alternative. Just reading the DASI protocol or listening to someone explaining what the DASI is about is insufficient. Different parts of the DASI, especially the part where physicians track detailed information concerning participation and activity limitations experienced by patients, need practice and feedback. This kind of practice takes time: At least three days are needed for instruction, role-playing and for feedback on actual patient interviews.
Refresher courses in the DASI method can be taken at the NSPOH and also at the Educational Branch of the Social Security Office ("UWV Opleidingen"). Because the DASI training has already been developed and field tested, almost no expenses are attached to developing the training program. Furthermore, there are no changes expected for the duration of the interview, or for patient satisfaction and disability pension. Moreover, the insurance physicians were satisfied with the training program. Because it is only a three-day course, expenses are limited.

In addition to the interview method, it would also seem useful to focus some attention on the guidelines and protocols available. Several guidelines now exist for insurance physicians, but we still found considerable differences among insurance physicians, for instance, when it came to deciding on hour limitations for work. It is possible that more agreement among physicians could be obtained if more attention was spent on the use of guidelines in daily practice. The creation of a guideline is often deliberate; after a literature search, a group of specialists develops a guideline which is commented upon by several insurance physicians as well as, for instance, patients and specialist physicians. However, research, supervision and evaluation in daily practice are less intensive than this and can be improved.

RECOMMENDATIONS

Future research

- We found that more functional limitations were assessed when the DASI was used as compared to usual practice. Therefore, further research into the validity of the DASI by comparing the outcome of DASI assessments with performance tests (FCE), questionnaires and other instruments would seem appropriate.
- We found that the content of the interview was important for the outcome. Comparing the outcome of DASI assessments with the outcome of other interview protocols should be conducted in order to check whether a different outcome between different interview protocols exists. If important differences are found to exist, then the need for one uniform interview protocol becomes that much more urgent.
- In the Netherlands, the insurance physician's interview plays an important role in assessments of functional limitations. Compared to other countries, the interview is deemed to be a valuable instrument[6], but other ways of assessing functional limitations should be explored. A combination of a performance test and/or questionnaires along with the usual assessment by an insurance physician, or assessments by more than one assessor, should be considered.
- We only studied the assessment of mental limitations in assessments based on written reports and video recordings. A real-life study would seem useful in order to study whether the results are also valid in real life.
- A relatively low agreement percentage among physicians was found on specific items of the FAL and MAL (e.g., "prolonged bending"), and almost no studies into the psychometric qualities of these instruments exist. Therefore, a study of the psychometric qualities and, if appropriate, an adaptation of the FAL and MAL would be recommended.
- We demonstrated that differences in collection of information, interpretation, and documentation can bring about inter-rater differences among physicians. In this thesis, we only studied interventions in information collection. Studies into interventions in interpretation (for instance, with a focus on guideline instructions), along with interventions in documenting

the functional limitations assessed in terms of inter-rater reliability, would offer an interesting perspective.

Practice

- A considerable difference among insurance physicians was found in assessing "hours a patient can work daily." This is undesirable because of the major consequences in terms of a disability benefit. The current "reduced working hours" guideline[15] is apparently inadequate or is not sufficiently applied by physicians. Therefore, we recommend creating a new guideline or coming up with a better implementation of the current one.
- We found that the same information could lead to different outcomes because of a different interpretation of this information. This was especially the case in patients who experience functional limitations, but where little or no objective medical findings are present. In our experience, the interpretation of functional limitations in diseases with few objective medical findings is a major point of discussion among insurance physicians in daily practice. The medical disability criterion in the Dutch "Assessment of Occupational Disability Decree"[19] is insufficiently conclusive to solve this problem. A decision has to be made as to what insurance physicians need to assess: the actual behavior of the patient, medical contra-indications given for a certain disease, or something in between. We think the professional group of insurance physicians, possibly in cooperation with politicians, needs to make a statement concerning this subject before creating new guidelines. Then, a conclusive guideline with anchor cases can be created. Mediprudence, as analogous to jurisprudence in the judicial system, is a collection of reports of well substantiated and weighted work-disability assessment decisions. They can be used as an anchor in cases which are difficult to interpret. The current development of "mediprudence" in the Netherlands[28] can play an important role in this matter.

Policy

- To improve univocal interpretation and documenting of assessed functional limitations we would recommend a clear, conclusive instruction course and checking on whether the guidelines are actually being implemented. This is important for the medical disability criterion and the "reduced working hours" guideline, but also in the interpretation of FAL items. It has to be clear whether functional limitations at the time of examination, or as they are expected to be at the time the disability benefit starts, need to be documented, and whether a FAL has to be filled in when there are functional limitations but no disease.
- It is common nowadays for nurse practitioners to back up physicians in carrying out well-defined tasks. Nurse practitioners support general practitioners and specialists in treating and monitoring patients with diabetes and rheumatoid arthritis, for instance. We found that nurse practitioners were able to help insurance physicians in collecting information by using the DASI as a semi-structured interview. Because a semi-structured interview protocol is used, the

tasks of the nurse practitioner are well defined, and the final assessment can remain in the hands of the insurance physician. The advantages in this are, for instance, that routine tasks can then be delegated, which makes work more interesting for insurance physicians, and saves on expenses.

- Considerable differences between insurance physicians exist in terms of, for instance, assigning hour limitations, even despite guidelines. We would recommend a pilot or field testing procedure before implementing the guidelines into practice. For instance, a "reduced working hours" guideline could be tested on physicians by showing them video recordings of patients, and then measuring whether they agreed in their assessments and what their comments on the guideline were. If only a limited degree of agreement exists, the guideline should not be implemented, because the goal of the guideline, to reach uniformity, cannot be attained. Furthermore, after implementing a guideline, supervision of implementation and evaluation in daily practice should be improved upon.

GENERAL CONCLUSION

We did not find any instruments in the literature with satisfactory psychometric qualities for assessing functional limitations in work disability claimants. The Disability Assessment Structured Interview (DASI) is a useful instrument in the Netherlands with acceptable reliability and an acceptable level of satisfaction registered by insurance physicians as well as patients. Concerning the validity of the DASI, we found satisfactory content validity, but further research into validity is recommended. To improve on the agreement among insurance physicians, further research into interpretation of information (guidelines and protocols) would seem to be more useful than research into data collection (performance tests, interview and questionnaires).

Reference List

1. Functional Ability List (FAL). Claim Beoordelings- en BorgingsSysteem (CBBS). [Claim Assessment and Monitoring System]. LISV: Amsterdam, 2002.
2. 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]. CTSV: Zoetermeer, 2000.
3. 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.
4. Spanjer J. Belastbaarheidsgerichte beoordeling, een onderzoek naar intra- en interbeoordelaars betrouwbaarheid van WAO beoordelingen. [Disability Assessment Structured Interview, a study of inter- and intra-rater reliability of disability pension assessments]. LISV: Amsterdam, 2000.
5. Boer WEL de, Wijers JHL, Spanjer J, van der Beijl I, Zuidam W, Venema A. Gespreksmodellen in de verzekeringsgeneeskunde. [Interview models in insurance medicine]. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2006;14:17-23.
6. De Boer WEL, Besseling JJM, Willems JHBM. Organisation of disability evaluation in 15 countries. Pratiques et Organisations de Soins 2007;38:205-17.
7. Wind H, Gouttebarge V, Kuijer PP, Sluiter JK, Frings-Dresen MHW. The utility of functional capacity evaluation: the opinion of physicians and other experts in the field of return to work and disability claims. International Archives Of Occupational And Environmental Health 2006;79(6):528-534.
8. International classification of functioning, disability and health. Geneva: World Health Organisation, 2001.
9. Huffelen JW. Wat verwacht u van de verzekeringsarts? … En, tevreden? [What do you expect of the insurance physician?]. Netherlands School of Public and Occupational Health, 2005.
10. Clark W, Haldeman S. The development of guideline factors for the evaluation of disability in neck and back injuries. Spine 1993;18:1736-1745.
11. Söderberg E, Alexanderson K. Sickness certification practices of physicians: a review of the literature. Scandinavian Journal of Public Health 2003; 31(6):460-474.
12. Lax MB, Manetti FA, Klein RA. Medical evaluation of work-related illness: evaluations by a treating occupational medicine specialist and by independent medical examiners compared. International Journal of Occupational & Environmental Health 2004;10:1-12.
13. Wolfson AM, Doctor JN, Burns SP. Clinician judgments of functional outcomes: how bias and perceived accuracy affect rating. Archives of Physical Medicine & Rehabilitation 2000;81:1567-1574.
14. Rainville J, Pransky G, Indahl A, Mayer EK. The physician as disability advisor for patients with musculoskeletal complaints. Spine 2005;30:2579-2584.
15. Standaard verminderde arbeidsduur. [Guideline reduced working hours]. Amsterdam: Lisv, 2007.
16. Boot CRL, Vercoulen JHMM, van der Gulden JWJ, Orbon KH, van den Hoogen H, Folgering HT. Sick leave in patients with obstructive lung disease is related to psychosocial and work variables rather than to FEV1. Respiratory Medicine 2005;99:1022-1031.
17. Swinkles RAHM, Dijkstra PU, Bouter LM. Reliability, validity and responsiveness of instruments to assess disabilities in personal care in patients with rheumatic disorders. A systematic review. Clinical And Experimental Rheumatology 2008;23:71-79.
18. De Bont AA, Meus C, Hazelaar G. Stoornissen, beperkingen en handicaps in de uitvoering. Stoornissen, beperkingen en handicaps in de uitvoering. [Impairments, limitations and handicaps in daily practice]. CTSV, Zoetermeer, 1998.

19. Schattingsbesluit. [Assessment Decree Occupational Disability Act]. Ministry of Social Affairs and Employment, Den Haag, 2004.

20. 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. Archives of Physical Medicine & Rehabilitation 2004;85:837-839.

21. Cohen J. A coefficient of agreement for nominal scales. Educational and psychological measurement 1960;20:37-46.

22. Sheskin D. Handbook of parametric and nonparametric statistical procedures. 3 ed. Taylor & Francis Ltd, 2004.

23. Popping R. On agreement indices for nominal data. In: Saris W.E., Gallhofer I.N. Sociometric Research. Volume 1. Data collection and scaling. 1988:90-105.

24. Schouten HJA. Measuring pairwise agreement among many observers. Biometrical Journal 1980;22:497-504.

25. Schouten HJA. Measuring pairwise agreement among many observers. II. Some improvements and additions. Biometrical Journal 1982;24:431-435.

26. Jekel JF, Katz DL, Elmore JG. Epidemiology, biostatistics, and preventive medicine. 2 ed. Saunders, 2001.

27. Brouwer S, Dijkstra PU, Gerrits EHJ, Schellekens JMH, Groothoff JW, Geertzen JHB et al. Intra- en interbeoordelaarsbetrouwbaarheid 'FIS-belastbaarheidspatroon'en 'Functionele Mogelijkheden Lijst'. [Intra- and inter-rater reliability 'Functional Information System' and 'Functional Ability List']. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2002;11:360-367.

28. Beoordelen, behandelen, begeleiden. Medisch handelen bij ziekteverzuim en arbeidsongeschiktheid. [Assessing, treating, guiding. Medical acting in sickness absence and disability]. Gezondheidsraad, Den Haag, 2005.

Summary

If workers are unable to work due to disability, they can apply for workers' compensation benefit. The disability benefit procedure begins with an assessment by an insurance physician of the patient's functional work limitations and abilities. Insurance physicians in the Netherlands base their opinions on information from both treating and occupational physicians, plus a physical examination and, most importantly, on the interview with the patient. The assessment of functional limitations will have major consequences. Therefore, it is imperative that different insurance physicians arrive at the same assessment (reliability), and that the proper functional limitations are assessed (validity). However, as far as the assessment of functional limitations by physicians in the Netherlands is concerned, no literature on this reliability and validity can be found.
In the Netherlands, the patient interview plays an important role in work disability assessment. Three interview models are described, and we have studied one of them, the Disability Assessment Structured Interview (DASI). The DASI is a semi-structured interview protocol with fixed topics which are largely based on the International Classification of Functioning, Disability and Health (ICF). The main topics are: introduction, work, impairments, the limitations to activity that are experienced, participation, the patient's opinion, and the physician's opinion. Two important characteristics of the DASI are its semi-structured way of interviewing the patient, and its method of inquiring about specific and detailed examples of limitations and of those concrete activities which the patient still can undertake.

In the first part of this thesis, we explored the sources of variation among physicians in their assessments of functional limitations, and we identified the instruments that can be used in the assessment of functional limitations. Moreover, we systematically researched the literature for instruments to assess claimants' functional limitations as found in workers' compensation claims, in addition to researching the reliability and validity of these assessment instruments.
In the second part of this thesis, we went on to study the characteristics of the DASI, the reliability and validity of the functional limitation assessments when using the DASI, and the opinion that insurance physicians have of the DASI.

Chapter 1 provides a general introduction, including the aim and an outline of this thesis. The research questions are:

1 What are the possible sources of variation in work disability assessment?
2 Which instruments are described that measure or assess functional limitations in claimants, and what are their psychometric qualities?
3 What effect does detailed information on functioning in addition to medical history-taking have upon the functional limitations assessed and on inter-rater reliability?
4 In their own opinion, are physicians able to assess functional limitations based on a written DASI report?
5 What are the characteristics of the DASI in daily practice?

6 What is the patient satisfaction evaluation for physicians who conduct a DASI interview?
7 What comments on the DASI do insurance physicians have?
8 What is the opinion about using the DASI that insurance physicians have?
9 What is the intra- and inter-rater reliability of functional limitations assessments using the DASI?
10 What is the content and concurrent validity of functional limitation assessments using the DASI?

Chapter 2 presents a model in which the possible sources of variation and the instruments used in work disability assessment are described. Inter-rater variability among physicians in assessing functional limitations may be caused by several sources, because several steps have to be made in the assessment and several instruments can be used. First, data have to be collected; the instruments that can be used for this are questionnaires, performance tests and medical assessments. Second, these data have to be interpreted; the instruments that can be used here are, for instance, guidelines and protocols. Then, the assessment has to be documented; in the Netherlands, an important instrument for this is the Functional Ability List (FAL). Finally, in all three of these steps the qualities of the assessor play an important role, qualities including the cultural background, norms and education of the assessor.

Chapter 3 presents a systematic review of the literature concerning instruments for assessing functional limitations in claimants for workers' compensation. Electronic literature searches of Medline, Embase, CINAHL and PsycINFO were performed to identify studies focusing on the psychometric properties of the instruments used to assess functional limitations in workers' compensation claimants. Two independent reviewers selected relevant articles and then evaluated the psychometric qualities of the instruments found. Four of those instruments were: the Roland-Morris Disability Questionnaire (RDQ), the Patient-Specific Functional Scale (PSFS), the Isernhagen Work System (IWS), and the Multiperspective Multidimensional Pain Assessment Protocol (MMPAP). The questionnaires (RDQ and PSFS) did not focus specifically on the work situation and measured functional limitations in a limited manner. The psychometric qualities of the IWS were poor to moderate. For the MMPAP, only the predictive validity was measured. The instruments assessed three to 34 physical functional limitations, and there were no instruments found that assessed mental limitations in claimants. In short, we were unable to find any instruments that had satisfactory psychometric qualities for assessing functional limitations in claimants.

Chapter 4 describes a study which examines the effect of detailed information about functioning (in addition to medical history-taking), on the functional limitations assessed, as well as on inter-rater reliability. Three different groups of nine insurance physicians each were given different kinds of information on 30 patients. The first group received only medical information, the second group received detailed information on the functioning of the patient, and the third group was provided with

both. The detailed information on functioning provided by the patient had no important influence on the inter-rater reliability of assessed functional limitations. However, there were significantly higher scores on assessed functional limitation items as compared to medical history-taking alone.

In **Chapter 5** we investigated whether physicians were able to perform a disability assessment based on a written report and what the inter-rater reliability of functional limitations between physicians was in these assessments. In total, 12 insurance physicians used written reports to assess functional limitations in 12 patients. The reports involved a DASI 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 the Functional Information System and the Mental Ability List. The quality of the reports was evaluated as reasonable to good. Half the physicians found the assessment based on the reports to be reasonably reliable, 25% found the opposite, and 25% were indecisive. The overall agreement among the insurance physicians was reasonable to good, with a mean agreement of 76% (range 64-88%). Agreement was low among the physicians concerning the number of hours a day a patient could function.

In **Chapter 6** we analyzed the content of DASI interviews, asked physicians for their comments on the DASI, and determined the inter- and intra-rater reliability of the assessments using the DASI. Five insurance physicians made 14 video recordings of DASI interviews. These video recordings were shown to 22 insurance physicians who were then asked for their comments on the interviews shown. The recordings were analyzed by measuring the duration of the different topics of the interviews. The inter-rater reliability among the 22 physicians was measured by computing the percentage agreement with respect to the mental and physical items of the Functional Information System (FIS) and the Mental Ability List (MAL).To measure the intra-rater reliability, the five insurance physicians who made the recordings were asked to fill out the FIS and MAL right after the recordings and after seeing the video again after six months.
The mean duration of the interviews was 33 minutes (range 19-77 min), and as much time on impairments was spent as on functional limitations and on activities (each comprised 30% of the whole interview). The mean duration of the introduction was 9% of the total interview, the perception of work 9%, the client's opinion on the functional limitations 8%, the decision of the physician 13%, and final reaction of the patient 4%. In general, the 22 physicians found that the interviews were structured, functional and efficient, and that in their own interviews more attention was given to medical issues and less to the functional limitations experienced by the patients. The mean percentage agreement among the 22 physicians on the items of the FIS and MAL was 74% (range 56-85%). There was a considerable difference among all physicians in assessing the hours a patient can work daily. The intra-rater percentage agreement was 80% (range 52-100%). When the physicians saw their own interviews on video after six months, they assessed fewer functional

limitations as compared to when they had just seen the patient.

Chapter 7 describes a study in which the reliability and validity of the DASI was investigated. A randomized controlled trial was conducted. Patients applying for a work-disability pension after 21 months of sick leave were independently interviewed and examined either by two physicians who had completed a DASI training period (n=32) or by two physicians from a control group (n=30) without any DASI training. Agreement percentages within both groups of physicians, eligibility for a disability benefit, and differences between the groups in terms of the scores given on the work-limitation items from the Functional Ability List (FAL) were measured to investigate reliability and concurrent validity. To determine the content validity, the insurance physicians who completed the DASI training (n=8) were asked to fill out a questionnaire concerning their opinion of the DASI. Additionally, patients were also asked to fill out a questionnaire to measure their satisfaction as to the behavioral aspects of the physicians.
The groups showed no important differences in agreement percentages; mean agreement percentage was 81% (range 59-91%). Content validity was satisfactory: all of the physicians found the DASI provided sufficient information in order to assess functional limitations and that it attended to all relevant aspects. The DASI was found to be an acceptable tool in daily practice, one that provided a realistic picture of the patient. Seven out of eight physicians found that the DASI allowed for a better basis for filling out the Functional Ability List than did their usual interview. In nine out of 21 items, the physicians of the control group indicated fewer work limitations compared to physicians using the DASI. However, this did not lead to an increased number of patients who qualified for a disability benefit. The patients' satisfaction report score for the physicians who conducted DASI interviews as well as in the control group was 7.7 on a scale from 1 to 10.

Chapter 8 summarizes and discusses the main findings of this thesis and comments upon the practical implications. Methodological issues are discussed and recommendations for further research, practice and policy are given. The main findings can be summarized by answering the research questions formulated in the first chapter:

1 Variation in assessing functional limitations may occur on three different levels: data collecting, interpretation and documentation of these data.
2 We did not find any instruments with satisfactory psychometric qualities for assessing functional limitations in work disability claimants.
3 Detailed information on functioning provided by the patient as compared to medical history-taking alone had no important influence on inter-rater reliability on assessed functional limitations. However, significantly higher scores on assessed functional limitation were found.
4 Half the physicians thought that a reliable assessment based on the written information was possible, 25% found the opposite, and 25% was indecisive.
5 The mean duration of the DASI interview without physical examination was 30-45 minutes.

About 30% of the total interview was spent on medical issues, and 30% on functional limitations and activities.

6 The patients' satisfaction report score for the physicians was 7.7 on a scale from 1 to 10. No difference was found between physicians who conducted DASI interviews and a control group.

7 Physicians who looked at video-recordings of DASI interviews indicated that the interviews were functional and efficient. However, they found that the medical anamnesis was too limited.

8 Physicians using the DASI in daily practice found that the DASI was an acceptable interview method that provided a realistic picture of the patient and provided sufficient information in order to assess functional limitations. The medical anamnesis was found to be sufficient too.

9 The intra- and inter-rater reliability in DASI assessments in general was moderate to good, but poor on the item "hours a patient can work daily." Inter-rater reliability was no better than with a control group that did not have any special training.

10 The content validity of the DASI was satisfactory. Physicians using the DASI indicated more serious functional limitations as compared to the usual practice without the DASI. Because there is no "gold standard" it is unknown who assessed the proper functional limitations.

The methodological issues discussed include the lack of a gold standard and the lack of sufficient psychometric properties for outcome measurements (for instance of the FAL).

The interview seems to be a valuable instrument; however, other ways of assessing functional limitations might be explored. A combination of a performance test and/or questionnaires and the usual assessment by an insurance physician or assessments by more than one assessor could be considered.

Inter-rater reliability seems to benefit more from a univocal interpretation of information, for instance, by administering proper guidelines and protocols, rather than by a univocally collecting of information in the interview. Therefore, it would seem useful to focus attention on guidelines and protocols, for example, on the "reduced working hours" guideline. We would recommend a pilot test or field testing before implementing any guidelines into practice. Furthermore, after implementing a guideline, surveillance of implementation and evaluation in daily practice could also be improved.

Although it would seem too early to use the DASI as a state-of-the-art interview protocol, because further research into its validity is needed, there are sufficient arguments present to offer DASI training as part of the education of insurance physicians and as a refresher course for insurance physicians already registered.

General conclusion

We did not find any instruments in the literature that had satisfactory psychometric qualities for assessing functional limitations in work disability claimants. The Disability Assessment Structured

Interview (DASI) is a useful instrument in the Netherlands with an acceptable reliability and along with satisfaction of insurance physicians as well as patients. Concerning the validity of the DASI, we found satisfactory content validity, but further research into validity is advisable. To improve agreement among insurance physicians, further research into the interpretation of information (guidelines and protocols) would seem to be more useful than research into data collection (performance tests, interview and questionnaires).

Samenvatting

Als werknemers door ziekte hun werk niet meer kunnen doen betaalt de werkgever het loon twee jaar door. Hierna kan een WIA uitkering bij de UWV worden aangevraagd. De beoordeling of de werknemer voor deze uitkering in aanmerking komt begint bij de verzekeringsarts, die de functionele beperkingen en mogelijkheden beoordeelt. Verzekeringsartsen in Nederland baseren hun oordeel voornamelijk op het gesprek dat ze met de cliënt hebben en daarnaast op het lichamelijk onderzoek en op de informatie van de bedrijfsarts en behandeld arts.

Deze belastbaarheidbeoordeling kan grote gevolgen hebben voor zowel de cliënt als de maatschappij. Het is daarom van belang dat verschillende verzekeringsartsen tot hetzelfde oordeel komen (betrouwbaarheid) en dat de juiste beperkingen en mogelijkheden worden vastgesteld (validiteit). In de literatuur zijn echter nauwelijks gegevens bekend over de betrouwbaarheid en validiteit van deze belastbaarheidbeoordelingen.

Het gesprek dat de verzekeringsarts heeft met de cliënt speelt een belangrijke rol bij de beoordeling van de belastbaarheid. In dit proefschrift wordt het Belastbaarheid Gericht Beoordelingsgesprek (BGB), nader onderzocht. De BGB is een semi-gestructureerd interview protocol met vaste gespreksonderwerpen, die grotendeels gebaseerd zijn op het model van de International Classification of Functioning, Disability and Health (ICF). De belangrijkste onderwerpen zijn: werk, stoornissen, ervaren belemmeringen, participatie, visie van de cliënt en de mening van de verzekeringsarts. Karakteristiek voor de BGB is de semi-gestructureerde manier van uitvragen en het doorvragen naar specifieke en gedetailleerde voorbeelden van beperkingen en concrete activiteiten die nog wel worden ondernomen.

In het eerste deel van dit proefschrift wordt in een model beschreven welke instrumenten gebruikt kunnen worden om de belastbaarheid te beoordelen en welke mogelijke oorzaken er zijn voor de variatie tussen artsen in belastbaarheidbeoordeling. Verder werd in de literatuur gezocht naar de betrouwbaarheid en validiteit van instrumenten om de beperkingen te kunnen beoordelen bij mensen die een uitkering aanvragen.

In het tweede deel worden vier studies beschreven waarin de betrouwbaarheid en validiteit werden onderzocht van belastbaarheidbeoordelingen, waarbij gebruik werd gemaakt van de BGB. Ook werd de mening van cliënten en verzekeringsartsen over de BGB gemeten.

Hoofdstuk 1 bevat een algemene introductie waarin het doel en de inhoud van dit proefschrift worden beschreven, uitmondend in de volgende onderzoeksvragen:

1. Wat zijn mogelijke bronnen van variatie in de beoordeling van belastbaarheid?
2. Welke instrumenten worden beschreven om de beperkingen te meten of beoordelen bij mensen die een uitkering aanvragen en wat zijn de psychometrische kwaliteiten?
3. Wat is, naast de medische anamnese, het effect van gedetailleerde informatie over functioneren op de beoordeling van beperkingen en de betrouwbaarheid?
4. Zijn artsen van mening dat ze beperkingen kunnen beoordelen op basis van een schriftelijk BGB rapport?

5. Wat zijn de kenmerken van de BGB in de dagelijkse praktijk?
6. Wat is de cliënttevredenheid over artsen die de BGB gebruiken?
7. Welk commentaar hebben verzekeringsartsen op de BGB?
8. Wat vinden verzekeringsartsen van de BGB als ze die zelf toepassen?
9. Wat is de intra- en interbeoordelaar betrouwbaarheid van beoordelingsgesprekken bij gebruik van de BGB?
10. Wat is de validiteit van beoordelingsgesprekken bij gebruik van de BGB?

In **Hoofdstuk 2** wordt in een model beschreven welke instrumenten gebruikt kunnen worden om de belastbaarheid te beoordelen en welke mogelijke oorzaken er zijn voor de variatie tussen artsen in hun oordeel over de belastbaarheid. Variatie tussen artsen in het beoordelen van belastbaarheid kan veroorzaakt worden op drie verschillende niveaus, met op elk niveau verschillende instrumenten die gebruikt kunnen worden. Ten eerste moeten gegevens worden verzameld. Instrumenten die hierbij gebruikt kunnen worden zijn vragenlijsten, functionele capaciteit evaluatie (FCE) en medische beoordelingen. Ten tweede dienen deze gegevens te worden geïnterpreteerd, waarvoor instrumenten als richtlijnen en protocollen gebruikt kunnen worden. Tenslotte dient het uiteindelijke oordeel te worden vastgelegd, waarvoor in Nederland vaak de Functionele Mogelijkheden Lijst (FML) wordt gebruikt. In alle drie stappen spelen persoonlijke eigenschappen van de beoordelaar een belangrijke rol, waaronder culturele achtergrond, normen en opleiding van de beoordelaar.

In **Hoofdstuk 3** wordt besproken dat er relatief veel onderzoek is verricht naar betrouwbaarheid en validiteit van instrumenten om de beperkingen te beoordelen bij patiënten in een revalidatie setting en minder bij mensen die een arbeidsongeschiktheidsuitkering aanvragen. Daarom werd een systematisch literatuuronderzoek verricht naar instrumenten om de beperkingen te beoordelen bij mensen die een uitkering aanvragen. Er werd gezocht in de bibliografische databestanden van Medline, Embase, CINAHL en PsycINFO naar studies die psychometrische kwaliteiten beschrijven van deze instrumenten. Twee beoordelaars selecteerden onafhankelijk van elkaar relevante artikelen en beoordeelden de psychometrische kwaliteiten van de gevonden instrumenten. Er werden vier instrumenten gevonden: de Roland-Morris Disability Questionnaire (RDQ), de Patient-Specific Functional Scale (PSFS), de Isernhagen Work System (IWS) en de Multiperspective Multidimensional Pain Assessment Protocol (MMPAP). De vragenlijsten (RDQ en PSFS) waren niet specifiek op de werksituatie georiënteerd en beperkingen werden slechts oppervlakkig gemeten. De psychometrische kwaliteiten van de IWS in de claimsituatie waren slecht tot matig. Van de MMPAP werd alleen de predictieve validiteit gemeten. De instrumenten beoordeelden drie tot 34 lichamelijke beperkingen en er werden geen instrumenten gevonden om mentale beperkingen te beoordelen. Er werden geen instrumenten gevonden met aangetoonde acceptabele psychometrische kwaliteiten.

In **Hoofdstuk 4** wordt een studie beschreven waarin cliënten gedetailleerde informatie over hun functioneren verstrekten. Het effect van deze informatie, naast de medische anamnese, op de beoordeelde beperkingen en de interbeoordelaar betrouwbaarheid werd onderzocht. Drie verschillende groepen van elk negen verzekeringsartsen kregen verschillende informatie over 30 cliënten: de eerste groep kreeg alleen medische informatie, de tweede groep kreeg gedetailleerde informatie over het functioneren van de cliënt en de derde groep kreeg beide.
De groep artsen die gedetailleerde informatie over het functioneren van de cliënt had gekregen concludeerde dat er zwaardere beperkingen waren dan de groep artsen die alleen van medische informatie was voorzien. De gedetailleerde informatie over het functioneren had geen duidelijke invloed op de interbeoordelaar betrouwbaarheid vergeleken met medische of beide soorten informatie.

In **Hoofdstuk 5** wordt een onderzoek beschreven waarin werd onderzocht of artsen op basis van een schriftelijk rapport een belastbaarheid beoordeling kunnen verrichten en wat in dat geval de interbeoordelaar betrouwbaarheid is. In totaal 12 artsen beoordeelden op basis van een schriftelijk rapport de belastbaarheid van 12 cliënten. De rapporten deden verslag van BGB gesprekken die gevoerd en vastgelegd waren door speciaal daarvoor opgeleide verpleegkundigen. De verzekeringsartsen werden gevraagd of ze van mening waren dat ze op basis van deze rapporten een betrouwbare beoordeling konden verrichten. Daarnaast werd de betrouwbaarheid tussen de artsen gemeten door het percentage overeenstemming op de items van het Functie Informatie Systeem (FIS) en de Psychische Mogelijkheden Lijst (PML) te berekenen. De kwaliteit van de rapporten werd door de artsen als redelijk tot goed beoordeeld. De helft van de artsen was van mening dat op basis van deze rapporten een redelijk betrouwbare beoordeling kon worden verricht, een kwart vond dat niet en een kwart had geen mening. De overeenkomst tussen de verzekeringsartsen was redelijk tot goed met een gemiddeld overeenstemmingspercentage van 76% (spreiding 64 – 88%). Wat betreft de urenbeperkingen (als de arts van mening is dat de cliënt geen hele dag kan functioneren) was de overeenstemming tussen de artsen laag.

In **Hoofdstuk 6** wordt een onderzoek beschreven waarin de inhoud van de BGB werd geanalyseerd, artsen naar hun commentaar op de BGB werd gevraagd en waarin de intra- en interbeoordelaars betrouwbaarheid van beoordelingen gebaseerd op de BGB werd bepaald. Totaal 14 BGB gesprekken van vijf verzekeringsartsen werden op de video opgenomen. Deze video opnames werden getoond aan 22 verzekeringsartsen, die hun commentaar gaven op de gesprekken. De gesprekken werden geanalyseerd door de duur van de verschillende gespreksonderwerpen te meten. De interbeoordelaar betrouwbaarheid tussen de 22 artsen werd gemeten door het percentage overeenstemming op de items van het Functie Informatie Systeem (FIS) en de Psychische Mogelijkheden Lijst (PML) te berekenen. De intra-beoordelaar betrouwbaarheid werd gemeten door de vijf verzekeringsartsen, die de opnames hadden gemaakt, te vragen de belastbaarheid te beoordelen vlak na de opname

en na het terugzien van de video na zes maanden.
De gemiddelde duur van de gesprekken was 33 minuten (spreiding 19 – 77 min). Er werd evenveel tijd besteed aan stoornissen als aan ervaren belemmeringen en participatie (beide 30% van het totale gesprek). De gemiddelde duur van de introductie was 9% van het totale gesprek, 9% ging over het werk, de visie van de cliënt duurde 8%, de mening van de verzekeringsarts 13% en de daarop volgende reactie van de cliënt duurde gemiddeld 4% van het hele gesprek. In het algemeen waren de artsen van mening dat de beoordelingsgesprekken gestructureerd, functioneel en efficiënt waren. In hun eigen beoordelingsgesprekken werd meer aandacht aan medische zaken besteed en minder aan de door de cliënten ervaren belemmeringen. Het gemiddelde overeenstemmingspercentage tussen de 22 artsen op de items van de FIS en PML was 74% (spreiding 56 – 85%).
Er was een aanzienlijk verschil tussen de artsen bij beoordeling van de urenbeperkingen. De intrabeoordelaar overeenkomst was 80% (spreiding 52 – 100%). Als de artsen hun eigen video opnames na 6 maanden weer terug zagen werden minder beperkingen gescoord in vergelijking met de beperkingen op basis van een zojuist verricht gesprek.

In **Hoofdstuk 7** wordt de betrouwbaarheid en validiteit van de BGB in een gerandomiseerd onderzoek met controlegroep onderzocht. Cliënten die na 21 maanden ziekte een WIA uitkering aanvroegen, werden geïnterviewd en onderzocht door twee artsen die een BGB training hadden ondergaan (n=32) of door twee artsen uit de controle groep zonder BGB training (n=30). De twee artsen beoordeelden de cliënten onafhankelijk van elkaar. Overeenstemmingspercentages, wel of niet in aanmerking komen voor een WIA uitkering en verschillen in scores op de FML items werden gemeten om de betrouwbaarheid en concurrente validiteit te onderzoeken. De inhoudsvaliditeit werd bepaald door de verzekeringsartsen, die een BGB training hadden ondergaan (n=8), een vragenlijst te laten invullen met betrekking tot hun mening over de BGB. Daarnaast vulden de cliënten een vragenlijst in om hun tevredenheid met betrekking tot gedragsaspecten van de artsen te meten.
Er werden geen duidelijke verschillen in overeenstemmingspercentages gevonden tussen de BGB en de controle groep, het gemiddelde overeenstemmingspercentage was 81 % (spreiding 59 - 91%). De inhoudsvaliditeit was goed: alle artsen vonden dat in de BGB aandacht wordt besteed aan alle relevante aspecten en dat de BGB voldoende informatie geeft om de beperkingen te beoordelen. De artsen waren van mening dat de BGB een realistisch beeld geeft van de cliënt en een acceptabele gespreksmethodiek is in de dagelijkse praktijk. Zeven van de acht verzekeringsartsen waren van mening dat de BGB een betere basis was om de FML in te vullen dan hun gebruikelijke beoordelingsgesprek. De artsen van de controlegroep gaven in 9 van de 21 FML items minder beperkingen aan dan de artsen van de BGB groep. Dit leidde echter niet tot meer uitkeringen in de BGB groep. De cliënttevredenheid met betrekking tot de artsen was in zowel de controle als in de BGB groep 7,7 op een schaal van 1 tot 10.

Hoofdstuk 8 geeft een samenvatting en discussie van de belangrijkste bevindingen van dit proefschrift. Verder worden de methodologische aspecten besproken en aanbevelingen gedaan voor verder onderzoek, de praktijk en beleid. De belangrijkste bevindingen kunnen worden samengevat op geleide van de 10 onderzoeksvragen:

1. Variatie in de beoordeling van belastbaarheid kan worden veroorzaakt op drie niveaus: gegevensverzameling, interpretatie en vastleggen van het oordeel.
2. In de literatuur werden geen instrumenten met voldoende psychometrische kwaliteiten gevonden om de belastbaarheid in een uitkeringssituatie te beoordelen.
3. Door de cliënt verstrekte gedetailleerde informatie over diens functioneren, naast de medische anamnese, had als gevolg dat artsen meer beperkingen aanwezig achtten. Deze informatie over het functioneren had geen invloed op de interbeoordelaar betrouwbaarheid.
4. De helft van de artsen was van mening dat een betrouwbare beoordeling verricht kan worden op basis van een schriftelijke rapportage, 25% vond dat niet en 25% had geen mening.
5. Een BGB gesprek zonder lichamelijk onderzoek duurde gemiddeld 30-45 minuten. Ongeveer 30% van het totale gesprek ging over medische zaken en 30% over beperkingen en activiteiten.
6. De cliënttevredenheid over artsen die de BGB gebruikten was met een score van 7,7 op een schaal van 1 tot 10 gelijk aan die van een controlegroep.
7. Verzekeringsartsen die video opnames van BGB interviews beoordeelden gaven aan dat de BGB functioneel en efficiënt was, doch dat de puur medische anamnese aan de magere kant was.
8. De artsen die de BGB zelf toepasten na een training waren van mening dat de BGB een realistisch beeld geeft van de beperkingen en mogelijkheden van de cliënt en een acceptabele gespreksmethodiek is in de dagelijkse praktijk. Ook de medische anamnese werd voldoende bevonden.
9. De intra- en interbeoordelaar betrouwbaarheid van beoordelingsgesprekken bij gebruik van de BGB was redelijk tot goed, behalve op het item "urenbeperking". De interbeoordelaar betrouwbaarheid was vergelijkbaar met een controlegroep zonder BGB training.
10. De inhoudsvaliditeit van de BGB was goed. Artsen die de BGB toepasten gaven meer beperkingen aan in hun beoordeling dan artsen zonder BGB training. Omdat een "gouden standaard" ontbreekt, is niet bekend wie de "juiste" beperkingen beoordeelde.

In de discussie over de methodologie worden het gebrek aan een gouden standaard en het gebrek aan voldoende psychometrische kwaliteiten van de uitkomstmaten, waaronder de FML, besproken. Het interview in het algemeen en de BGB in het bijzonder lijkt een waardevol instrument, maar onderzocht zou moeten worden of aanvullende instrumenten een rol in de beoordeling van functionele mogelijkheden kunnen spelen. Zo kan een combinatie van functionele capaciteittesten en/of vragenlijsten naast het onderzoek van de verzekeringsarts overwogen worden; of

beoordelingen door meerdere verzekeringsartsen van dezelfde cliënt.
Om de variatie tussen verzekeringsartsen te reduceren lijkt het eenduidig interpreteren van informatie meer bij te dragen dan eenduidige gegevens verzameling in het beoordelingsgesprek. Het is daarom nuttig aandacht te besteden aan onderzoek naar - en ontwikkeling van duidelijke richtlijnen en protocollen, bijvoorbeeld de standaard verminderde arbeidsduur. Aanbevolen wordt om richtlijnen en protocollen eerst in de praktijk uit te testen alvorens ze te implementeren in de dagelijkse praktijk. Verder kunnen de controle op de uitvoering en evaluatie na implementatie nog verbeterd worden.
Het is te vroeg om te adviseren de BGB landelijk als voorkeurs gespreksmethodiek in te voeren, omdat het wenselijk is eerst verder onderzoek naar de validiteit van de BGB te verrichten. Wel zijn er voldoende argumenten om de BGB training aan te bieden aan verzekeringsartsen in opleiding en aan geregistreerde verzekeringsartsen in het kader van nascholing.

Algemene conclusie

In de literatuur werden geen instrumenten met voldoende psychometrische kwaliteiten gevonden om de belastbaarheid in een uitkeringssituatie te beoordelen. Het Belastbaarheid Gerichte Beoordelingsgesprek (BGB) lijkt dit manco op te kunnen vullen, omdat het een bruikbaar instrument is met acceptabele betrouwbaarheid en tevredenheid bij zowel verzekeringsartsen als cliënten. Er werd een goede inhoudsvaliditeit aangetoond, echter verder onderzoek naar de validiteit is nodig omdat verzekeringsartsen die de BGB toepasten meer beperkingen aangaven dan een controlegroep die de BGB niet toepaste. Om de overeenkomst in beoordelingen tussen verzekeringsartsen te verbeteren lijkt onderzoek naar interpretatie van informatie (richtlijnen en protocollen) meer zin te hebben dan verder onderzoek naar de gegevensverzameling (FCE, interview en vragenlijsten).

Addendum 1

Functional Abilities List (FAL)

FUNCTIONAL ABILITIES LIST

*This list is an overview of an individual's **general** functional abilities during a full working day (minimum of 8 hours). **Restrictions** to these abilities with regard to **normal values** are given in a separate list if considered symptomatic of an illness, incapacity or accident in the opinion of the insurance company doctor. The standard functional levels required in daily life have been taken as the normal values. Unless expressly stated otherwise, incidental **peak demands** above the given functional levels are also possible.*
*This list should only be applied if accompanied by a **medical insurance physician's report** that, based on an analysis of the problem, evaluates, motivates and describes the correlation of functional abilities and limitations.*

Name:.....
National insurance number:..... .....sex***: m/f***
Diagnosis code:.....
Last/current work:.....
(hours per week:**.....**)
Resumed work:.....
(hours per week:**.....**)

Conclusion:

O **The client has long-term capacities for work**
O **The client does not have long-term capacities for work**

Explanation:

O **The client is capable of fully functioning in his own job/function**
O **The client is capable of functioning normally (see headings)**
O **The client's normal functioning is impaired (see headings)**
O **Other, see report by the insurance company doctor**
O **The client's personal and/or social functioning is extremely limited (see headings I, II)**
O **The client has been admitted to a hospital or institution recognised by the Exceptional Medical Expenses Compensation Act (AWBZ)**
O **The client is bedbound (for most of the day, long-term)**
O **The client is highly dependent as regards performing daily living activities (*ADL*)**
O **The client has highly variable functional abilities/loss of functional abilities**
< 3 months - 1 year

Date: **Insurance physician :**.....

HEADING I: PERSONAL FUNCTIONING

1. Focusing attention

0 Normal, can concentrate on an information source (book, documentary on TV or radio) for at least half an hour)
1 Limited, cannot concentrate on an information source (newspaper, current affairs programme on radio or TV) for more than half an hour
2 Very limited, cannot concentrate on an information source (advertising brochure, TV or radio advert) for longer than 5 minutes

2. Dividing attention

0 normal, can concentrate for at least half on a number of information sources (can manage driving or cycling in busy traffic)
1 limited, cannot concentrate for at least half on a number of information sources (can manage driving or cycling in busy traffic)
2 Very limited, cannot concentrate for longer than 5 minutes on a number of information sources (crossing a busy street alone)

3. Memory

0 normal, can generally remember relevant things promptly, without resorting to unusual aids
1 limited, must frequently write things down as a memory aid to safeguard the continuity of his actions
2 very limited, constantly unable to remember essential everyday things (time, place, person, subject), and cannot compensate with memory aids

4. Insight into own abilities

0 normal, mostly estimates own abilities and limitations reasonably accurately
1 limited, generally highly overestimates own abilities
2 limited, generally highly overestimates own limitations

5. Effective action (task implementation)

(coordinated action, gears own activities to realising a goal)

0 normal, no specific limitations to his effective action. The routine of daily life (getting up on time, washing, dressing, preparing breakfast, breakfasting, locking up the house and arriving at appointments on time)
1 limited, does not commence activities on time in order to realise set goal
2 limited, does not conduct the necessary activities in a logical order
3 limited, does not check the course of the activities
4 limited, does not end the activities once set goal is reached or cannot be reached
5 otherwise limited in taking effective action, i.e.....

6. Independent action (carrying out tasks autonomously)

0 normal, no specific limitations to independent action in daily life
1 limited, does not generally initiate action
2 limited, does not generally set himself goals
3 limited, does not generally think of variations on a task independently
4 limited, generally does not generally take an independent decision on the best approach to take
5 limited, does not generally realise when the decided approach falls short
6 limited, in those instances, does not generally take an independent decision to follow an alternative line of action or set a different goal
7 limited, does not generally continue, under own initiative, until goal is accomplished
8 limited, does not call on others promptly for help when the situation demands
9 otherwise limited in independent action, i.e.....

7. Action tempo

0 normal, there are no specific limitations to the action tempo in daily life
1 limited, the action tempo is considerably slower

8. Other limitations to personal functioning

0 normal, no other specific limitations to personal functioning in daily life
1 limited, other specific limitations, i.e.....

9. Specific conditions for personal functioning in a work situation

(is work functioning dependent on specific conditions because of the said limitations or the client's compensatory behaviour?)

0 no, there are no specific conditions for personal work functioning
1 yes, the client has been advised to follow a fully pre-structured work schedule: concrete, one-sided assignments (what, when, how long; one task per assignment) and to follow prescribed implementation orders (how)
2 yes, the client has been advised to follow fixed, familiar working methods (routine-dependent)
3 yes, the client has been advised to perform work under immediate supervision (consistent feedback) and/or to work under intensive supervision
4 yes, the client has been advised to work in a situation in which he is not distracted by the activities of others
5 yes, the client has been advised to work in a predictable working situation, cannot respond flexibly to highly varied situations in which in work s performed and/ or varied work content
6 yes, the client has been advised to work in a work situation not susceptible to constant interruptions and disturbance
7 yes, the client has been advised to work in a work situation not susceptible to constant deadlines or production peaks
8 yes, the client has been advised to work in a work situation in which a high action tempo is not required
9 yes, the client has been advised to work in a work situation in which there is no increased personal risk
10 yes, there are other specific conditions, i.e.....

Explanation: *see medical insurance physician's report*

HEADING II: SOCIAL FUNCTIONING

1. Vision

o normal, no specific limitation in daily functioning
1 limited, i.e.....

2. Hearing

o normal, no specific limitation in daily functioning
1 limited, i.e.....

3. Speech

o normal, no specific limitation in daily functioning
1 limited, i.e.....

4. Writing

0 normal, no specific limitation in daily functioning
1 limited, i.e.....

5. Reading

0 normal, no specific limitation in daily functioning
1 limited, i.e.....

6. Dealing with the emotional problems of others

0 normal, can generally empathise with the problems of others but can also distance himself in terms of behaviour and experience
1 limited, generally becomes involves in the problems of others; nevertheless, can distance himself sufficiently in terms of behaviour although not experience
2 very limited, generally identifies with the problems of others and cannot distance himself in terms of either behaviour or experience

7. Expressing personal feelings

0 normal, can generally express personal feelings in a way acceptable to others, both verbally and behaviourally
1 limited, confuses others with unpredictable or unconventional ways of expressing feelings
2 very limited, is generally incapable of expressing feelings (blocks himself) or expresses them in an uncontrolled way regardless of the feelings of others

8. Dealing with conflicts

0 normal, can directly deal with conflicts with aggressive or unreasonable people
1 limited, can only deal with conflicts with aggressive or unreasonable people by phone or in writing
2 very limited, cannot generally deal with conflicts

9. Working with others

0 normal, can jointly carry out a task with others (teamwork)
1 limited, can work with others but only with a task of his own, clearly defined beforehand
2 very limited, as a rule is unable to work with others

10. Transportation

0 normal, can drive or cycle or use public transport on his own
1 limited, is reliant on others for transportation

11. Other limitations to social functioning

0 normal, no other specific limitations to social functioning in daily life
1 limited, other specific limitations, i.e. .....

12. Specific conditions for social functioning at work

(is social functioning at work dependent on specific conditions because of the said limitations or the client's compensatory behaviour?)

0 no, there are no specific conditions for social functioning at work
1 yes, the client has been advised to work in a situation demanding no direct contact with clients (some occupations in the service sector)
2 yes, the client has been advised to work in a situation where little or no direct contact with patients or those needing help is required (some occupations in the health care sector)
3 yes, the client has been advised to work in a situation in which, if necessary, he can fall back on immediate colleagues or managers (no solitary job)
4 yes, the client has been advised to work in a situation which generally does not require direct contact with colleagues
5 yes, the client has been advised to work in a situation involving no managerial aspects
6 yes, there are other specific conditions, i.e.....

Explanation: *see medical insurance physician's report*

HEADING III: ADJUSTING TO PHYSICAL ENVIRONMENT

1. Heat

0 normal, no specific limitations
1 limited, i.e. .....

2. Cold

0 normal, no specific limitations
1 limited, i.e.....

3. Draught

0 normal, no specific limitations
1 limited, i.e.....

4. Skin contact

0 normal, no specific limitations
1 limited, i.e.....

5. Protective measures

0 normal, no specific limitations
1 limited, i.e.....

6. Dust, smoke, gases and fumes

0 normal, no specific limitations
1 limited, i.e.....

7. Noise nuisance

0 normal, no specific limitations
1 limited, i.e.....

8. Vibration

0 normal, no specific limitations
1 limited, i.e.....

9. Other limitations to physical adjustment abilities

0 normal, no other specific limitations to physical adjustment abilities
1 allergies, i.e. .....
2 increased susceptibility to infections, i.e.....
3 weakened skin barrier, i.e. .....
4 other limitations, i.e. .....

10. Specific conditions for adapting to the physical working environment

(is adjustment to the working environment dependent on specific conditions because of the said limitations or the client's compensatory behaviour?)

0 no, there are no specific conditions for adapting to the physical working environment
1 yes, there are specific conditions for adapting to the physical working environment, i.e.....

Explanation: *see medical insurance physician's report*

HEADING IV: DYNAMIC MOVEMENT

1. Dominance

0 not applicable
1 right
2 left

2. Localisation limitations

0 neither right nor left
1 right
2 left
3 both sides

3. Use of hand and fingers

0 normal, no specific limitations when using hands and fingers in daily life
1 limited, can hardly perform a ball grip, if at all
2 limited, can hardly perform a pen grip, if at all
3 limited, can perform a pincer grip, if at all
4 limited, can perform a key grip, if at all
5 limited, can perform a cylinder grip, if at all
6 limited, can use hand/fingers to squeeze or grip , if at all
7 limited, is hardly able to perform fine motor hand/finger movements
8 limited, is not able to perform repetitive hand/finger movements, if at all

4. Touch

0 normal, no specific limitations in daily life
1 limited, i.e.....

5. Using a keyboard and mouse

0 normal, can perform all required movements
1 limited, i.e.....

6. Working with a keyboard and mouse

0 normal, if required can use a keyboard and mouse most of the working day (professional word-processing, programming, CAD/CAM work, electronic sales)
1 slightly limited, if required can use a keyboard and mouse half the working day (roughly 4 hours) (policy worker)
2 limited, if required can use a keyboard and most for a small part of the working day (roughly 1 hour) (to send email)
3 very limited, can use a keyboard and mouse less than thirty minutes a working day

7. Twisting movement – hand and arm

0 normal, no specific limitations in daily life
1 limited, i.e.....

8. Stretching arm

0 normal, can stretch arms (serve coffee)
1 slightly limited, can stretch arm slightly (shoulder-hand distance = 50-60 cm)
2 limited, can stretch arm slightly (shoulder-hand distance = less than 50 cm)

9. Can stretch arm frequently during work (roughly 20 times a minute)

0 normal, if required can stretch frequently during each hour of the working day (cashier work in wholesale company, packaging work)
1 slightly limited, if required can stretch frequently for roughly 4 hours of the working day
2 limited, if required can stretch frequently roughly one hour per working day
3 very limited, cannot stretch frequently during one hour of the working day

10. Bending

0 normal, can bend roughly 90 degrees (pick up a piece of paper from the ground)
1 limited, can bend roughly 60 degrees (pick up a bag from the ground)
2 very limited, can bend roughly 45 degrees (pick up crumbs from a chair)

11. Frequent bending during work (roughly ten times per minute)

0 normal, if required, can bend frequently during each hour of the working day
1 slightly limited, if required, can bend frequently roughly 4 hours per working day
2 limited, if required, can bend frequently one hour per working day
3 very limited, cannot bend frequently one hour per working day

12. Turning/twisting

0 normal, can turn torso at least 45 degrees (look behind while cycling, reach into the back seat of the car to get a bag while sitting in the front)
1 limited, i.e.....

13. Pushing/pulling

0 normal, can push or pull roughly 15 kgf (remove a stubborn cork from a wine bottle
1 limited, can push or pull roughly 10 kgf (full rubbish container)
2 very limited, can push or pull roughly 5 kgf (open door with door-closer)

14. Carrying/lifting

0 normal, can carry roughly 15 kg (toddler)
1 slightly limited, can carry roughly 10 kg (infant)
2 limited, can carry roughly 5 kg (bag of potatoes)
3 very limited, can lift roughly 1 kg (litre container of milk)

15. Frequently managing light objects at work (roughly 10 times per hour)

0 normal, if required can manage objects weighing around 1kg frequently during every hour of the working day (order book)
1 slightly limited, if required, can manage objects of 1 kg for roughly 4 hours per working day
2 limited, if required, can manage objects of around 1 kg for roughly one hour per working day
3 very limited, cannot manage objects of around 1 kg for one hour per working day

16. Frequently managing heavy loads at work (roughly 10 times per hour)

0 normal, if required, can frequently manage loads of roughly 15 kg for one hour per working day
1 limited, cannot frequently manage loads of roughly 15 kg during one hour per working day

17. Head movements

0 normal, can move head without hindrance
1 limited, can move head to a limited extent
2 very limited, can barely turn head to the side if at all
3 very limited, can barely move head up and down if at all

18. Walking

0 normal, can walk for roughly one consecutive hour (a walk)
1 slightly limited, can walk for roughly 15-30 consecutive minutes (a stroll)
2 limited, walk for roughly 5-15 consecutive minutes (to the letterbox)
3 very limited, can walk for less than 5 consecutive minutes (indoors)

19. Walking while at work

0 normal, if required, can spend most of the working day walking (postal worker)
1 slightly limited, if required can walk half the working day (roughly 4 hours)
2 limited, if required, can walk a limited part of the day (roughly 1 hour)
3 very limited, can walk for less than half an hour per working day

20. Stairclimbing

0 normal, can walk at least 2 flights of stairs up and down in one go (2 floors of a house)

1 slightly limited, can walk at least up and down a flight of stairs in one go (1 floor of a house)
2 limited, can walk at least down stairs in one go (1 floor of a house)
3 very limited, can only walk on or off the curb in one go

21. Climbing
0 normal, can at least climb up and down a ladder (1 floor)
1 slightly limited, can at least climb up and down a household stepladder
2 limited, can at least get up and down from a stool (50 cm, elephant foot)
3 very limited, cannot climb up and down

22. Kneeling or squatting
0 normal, can touch the ground with hands when kneeling or squatting (picking up a coin)
1 limited, can barely touch the ground with hands when kneeling or squatting, if at all

23. Other limitations to dynamic movement
0 normal, no other specific limitations to dynamic movement in daily life
1 other specific limitations, i.e.....

24. Specific conditions for dynamic movement at work
(is dynamic movement at work dependent on specific conditions because of the said limitations or the client's compensatory behaviour?)
0 no, there are no specific conditions for dynamic movement at work
1 yes, there are specific conditions for dynamic movement at work, i.e.....

Explanation: *see medical insurance physician's report*

HEADING V: STATIC MOVEMENTS

1. Sitting
0 normal, can sit for roughly 2 consecutive hours (car journey)
1 slightly limited, can sit for roughly one consecutive hour (film)
2 limited, can sit for roughly 30 consecutive minutes (meal)
3 very limited, can sit for less than 15 consecutive minutes (TV news)

2. Sitting at work
0 normal, if required, can sit for almost the whole working day (assembly work, cashier work, administrative work)
1 slightly limited, if required can sit for most of the working day (6-8 hours)
2 limited, if required can sit for half the working day (roughly 4 hours)
3 very limited, can sit for less than 4 hours per working day

3. Standing
0 normal, can stand for roughly 1 consecutive hour (spectator at sports events)
1 slightly limited, can stand for roughly 15-30 consecutive minutes (waiting in line for theme park attraction)
2 limited, can stand for roughly 5-15 consecutive minutes (washing up)
3 very limited, can stand for less than 5 consecutive minutes (brushing teeth)

4. Standing during work
0 normal, if required, can stand for almost the whole working day (sales jobs, production line jobs)
1 slightly limited, if required can stand for half the working day (roughly 4 hours)
2 limited, if required can stand for a limited part of the working day (roughly 1 hour)
3 very limited, can stand for less than 30 minutes per working day

5. Active kneeling or squatting

0 normal, can perform activities kneeling or squatting for at least 5 minutes (gardening)
1 limited, can perform activities for less than 5 consecutive minutes (cleaning kitchen cupboard door)

6. Active bending and/or twisting

0 normal, can perform activities bending or twisting for at least 5 minutes (sweeping steps)
1 limited, can perform activities bending or twisting for less than 5 consecutive minutes (tying shoelaces)

7. Active above shoulder level

0 normal, can perform activities above shoulder level for at least 5 minutes (hanging up curtains)
1 limited, can perform activities bending or twisting for less than 5 consecutive minutes (changing a light bulb)

8. Keeping head in a certain position during work

0 normal, if required, can keep head in a certain position for almost the whole working day (screen work, quality control)
1 slightly limited, if required, can keep head in a certain position for half of the working day (roughly 4 hours)
2 limited, if required, can keep head in a certain position for a limited part of the working day (roughly 1 hour)
3 very limited, can keep head in a certain position for less than thirty minutes per working day

9. Changing position

0 normal, no specific sequence of different positions required
1 specific requirements of various positions required, i.e.....

10. Other limitations to static movement

0 normal, no other specific limitations in daily life
1 other specific limitations, i.e.....

11. Specific conditions for static movement at work

(are static movements at work dependent on specific conditions because of the said limitations or the client's compensatory behaviour?)

0 no, there are no specific conditions for static movements at work
1 yes, there are specific conditions static movements at work, i.e.....

Explanation: *see medical insurance physician's report*

HEADING VI: WORKING HOURS

1. Periods in a day (24 hours)

0 normal, if required can work at any hour of the day, night included
1 limited, cannot work nights (00.00 - 06.00)
2 limited, cannot work evenings (18.00 - 24.00)

2. Hours per day

0 normal, can work at least 8 hours per day
1 somewhat limited, cannot work on average more than 8 hours per day
2 slightly limited, cannot work on average more than roughly 6 hours per day
3 limited, cannot work on average more than roughly 4 hours per day
4 extremely limited, cannot on average work more than roughly 2 hours per day

3. Hours per week

0 normal, can work an average of at least 40 hours per week
1 somewhat limited, can work an average of roughly 40 hours per week
2 slightly limited, can work an average of roughly 30 hours per week
3 limited, can work an average of roughly 20 hours per week
4 extremely limited, can work an average of roughly 10 hours per week

4. Other limitations with regard to working hours

0 normal, there are no other specific limitations regarding working hours
1 other specific limitations, i.e.....

Explanation: *see medical insurance physician's report*

Addendum 2

Disability Assessment Structured Interview (DASI)

Version 5.2 (October 2007)

Jerry Spanjer

ABSTRACT

The Disability Assessment Structured Interview (DASI) is a semi-structured method for collecting data to assess work limitations and its prognosis. The emphasis is on a semi-structured interview with the patient. Other data such as observation, physical examination and information from treating physicians also play a role in the assessment. In the interview, the three levels of functioning – impairments, activity restrictions and participation – are mapped in a structured way in accordance with the 'International Classification of Functioning, Disability and Health' (ICF)[1]. An important component of the interview is inquiring about specific and detailed examples of restrictions and of activities which the patient is still undertaking. Work limitations can be assessed on the basis of the collected data.

DISABILITY ASSESSMENT STRUCTURED INTERVIEW

A semi-structured interview for assessment of work limitations

INTRODUCTION

In both Europe and the United States, it is doctors who assess a patient's fitness for work. The Netherlands is unique in that it works with an interview protocol.[2] However, assessing whether a patient is able to work is not part of a physician's training. The basic physician's training teaches physicians to use a set format, in which further tests such as laboratory tests and X-rays follow the anamnesis and examination. The diagnosis is made on the basis of this information and a therapy can be initiated on the basis of the diagnosis.

The main task of insurance physicians is not to treat people by means of therapy, but to assess their capacity for work and the prognosis of that capacity. Because the diagnosis does not say anything about the nature and severity of the disability, a different methodology is required than what is taught as part of the basic physician's training. For example, a diagnosis of rheumatoid arthritis says nothing about the nature and severity of a patient's disability.[3] And with COPD, the parameters indicating the severity of the illness, such as the degree of lung obstruction (FEV1), are not associated with the severity of restrictions.[4] Another difference between curative and insurance physicians is that the former deal with patients who benefit from a correct diagnosis, while the latter deal with patients who do not necessarily benefit from an accurate assessment of their capacity for work. For example, a patient may paint things as being less rosy than they actually are for financial reasons, or because of the implications in terms of employment laws.

Because the insurance physician's assessment is largely based on the interview with the patient, and because standardized data gathering is essential for a low inter-doctor variation,[5] it is important that interviews have a semi-structured format. At present, there are three interview methods in

the Netherlands with some form of protocol, one of which – the Disability Assessment Structured Interview (DASI) – I will discuss here.[6]

DASI is a semi-structured method for gathering data enabling the physician to arrive at an assessment of a patient's restrictions and work capacity. The data gathered by the physician may vary considerably, for example oral information from the patient, physical and psychological examinations, information from the physician or employer, and hetero-anamnesis. Additional tests may also be carried out, for example expert opinion from a specialist, psychological tests, performance tests (also called functional capacity evaluation or FCE), and questionnaires. With DASI, the emphasis is on a semi-structured interview with the patient.
The physician makes a substantiated assessment of the patient's capacity for work based on all the information collected. DASI simply describes what information should be collected and how; it does not indicate how the assessment itself should be arrived at. This is outlined inter alia in the Dutch Disability Benefits Act,[7] and in health-insurance guidelines, standards and protocols. Other influences include current medical practice (e.g. 'for back complaints, we advise movement rather than bed rest') and peer testing. However, this approach still leaves many grey areas, for example how to deal with complaints that are difficult to objectify. Perhaps new instruments, such as medical jurisprudence (assessment examples, like jurisprudence in legal judgements), can still play a role in reducing inter-assessor variation. Be that as it may, it is still important for assessors to base their decisions as much as possible on the same information – hence the importance of semi-structured assessment interviews.

CONTENT AND FORM

The Dutch Disability Benefits Act states that impairments (functioning at the physical and psychological levels), activity restrictions (at the level of behaviour and activities) and handicaps (at the level of social roles) must be logical and coherent, and there must be consistency. In the ICF, this corresponds to the following levels of functioning: impairments, activity limitations and participation. These three levels are explicit and systematically enquired about in the DASI.

- Impairments: these are mapped by asking for medical data and complaints, and by a physical and psychological examination.
- Activity limitations: structured mapping of specific, detailed examples of the activity restrictions experienced by the patient.
- Participation: structured mapping of what the patient can and can no longer do in the household, sports, hobbies, social contact and work. A detailed description of a normal day is also asked for.

In the legal literature, recommendations are made to assess whether an interview is credible.[8,9] An

important component is the analysis of interview content – the Criteria-based Content Analysis (CBCA).[10,11] The CBCA provides criteria to assess whether an interview is plausible. It assumes that a person who has experienced something will make a different statement from someone who has not. Only those who have experienced something themselves can meet the CBCA criteria. General characteristics include a logical structure, the unstructured production of a statement and the amount of detail.

Characteristics of DASI are therefore:

- All topics must be covered – a logical structure means that a statement should contain no contradictions or inconsistencies. To assess this, a statement must be made on every topic.
- Structured questioning – it is extremely difficult to make consistent statements in a disorganized way without genuine experience. It is therefore important to conduct the interview in a fixed, non-coherent way. This compels the patient to produce statements in an unstructured way. Consistency can be examined by assessing whether the patient provides the same information for the fields of impairments, activity limitations and participation. For example, if patients say that they cannot lift a heavy object, they should also mention this in relation to housework or their description of a normal day.
- Questioning for specific details – by making things specific and asking for details, the assessor makes simulation or dissimulation difficult for patients. Consistency between parts of a statement can also be examined in more detail. Variation between assessors is then likely to decrease because there is less room for interpretation. For instance, if a patient claims to have a headache 'regularly', there is more scope for interpretation than if the patient says 'once a month'. This will be discussed in more detail in Appendix 1.

The DASI is a semi-structured interview protocol in which the subject matter and content is fixed but not the questions that patients are asked. It differs from a questionnaire in that the assessor can communicate at the patient's level or take the patient's emotional handicaps into consideration. It is not always easy to adhere to a fixed structure or to obtain concrete examples. There are several reasons for this. For example, the patient may want to talk about things other than what the assessor wants to hear or may have difficulty adhering to a structure.

Furthermore, the assessor himself can disrupt the interview structure when he tends to give advice and try to influence the patient's behaviour during the interview ('If you are an interviewer, you are not a therapist'). The assessor's communication skills are therefore an important tool. It is important to give patients a chance to tell their story without giving advice and making value judgements during their statement.[12] If there are clear inconsistencies in the statement, the patient can be asked for an explanation at the end of the interview.

The DASI is a method of collecting relevant data to assess work limitations and its prognosis. In addition to interviewing, other important data-collecting methods are patient observation, physical

examination and collecting information from treating physicians. Laboratory or imaging studies and other tests may also be necessary. The physician has to take all this information into consideration and assess work limitations. He or she must assess whether the information is consistent and plausible and whether the behaviour the patient presents is in line with the medical situation. Different physicians may have different interpretations, which is why other instruments such as interpretation guidelines are required in addition to a structured anamnesis.

Research into inter-rater variation for physicians using the DASI showed an acceptable reliability.[13,14,15] Research into the validity of the DASI method is difficult because there is no golden standard. Because the DASI is based on the ICF levels, the method's content validity seems satisfactory.

Disability Assessment Structured Interview

A. Introduction

- Introduction
- Putting the patient at ease
- Short explanation of the procedure
- Summarize the known data

B. Work

- Type and duration of work
- Work content
- Perception of work

C. Impairments

1. Medical data:

- Medical history and nature of current complaints
- General anamnesis, other diseases
- Course of disease
- Cause of disease (private/personal/work (PPW), diagnosis of treating physician)
- Treatment (now and in the past, medication, physician's opinion)

2. Examination:

- Physical and psychological
- If necessary, information from treating physician, heteroanamnesis, expert consultation

D. Activity limitations

- Nature of experienced activity limitations
- Seriousness (specific, detailed examples)

E. Participation

- Which activities does the patient do and which activities does he/she not (or no longer) do? (description of activities during a normal day, ADL, sports, hobbies, housekeeping, work, personal and social functioning)

F. Patient's view

- What is the patient's view of his or her capacity to function?
- What is the patient's response to the provisional opinion of the physician?

G. Physician's opinion

- The physician communicates his or her judgement and its possible consequences
- Explanation of the rest of the procedure
- Opportunity for the patient to respond and rounding off

Work

In addition to the contents of the patient's job and the number of hours worked a week, it is important to know how long the patient has worked in this job with this employer. The patient is also asked how he or she experienced his or her work. Key information here is which aspects of the job were considered as heavy and what the work relationships were like.

why This shows whether or not the patient is motivated to go to work. The assessor can investigate whether not being able to perform the job (or parts of the job) is consistent with the restrictions. If the patient has worked for a short time at this job, he or she may have been incapable of doing it from the outset.

example A patient with light administrative duties has called in sick with mild back complaints. There have been many problems with colleagues and the boss. This seems to be a motivational rather than a health problem.

Medical data

Medical history and the nature of the complaints:

Mapping the history of the present disease and medical anamnesis – this is what doctors learn in their medical training. A single complaint is easier than complex diseases and multiple complaints, which can lead to an impenetrable mass of complaints, diagnoses, therapies and patient and therapists' opinions. Providing structure can help to elicit the desired information. An inventory can be made firstly of the groups of complaints that the patient is experiencing, and secondly of the complaints within each group. Each complaint can then be attended to. In this way, an inventory can be compiled within a relatively short time without missing any complaints.

example A patient with fibromyalgia has two groups of complaints – pain and tiredness. The pain is in the knee, neck, shoulders and back. First the patient is asked: where do you feel the pain? The patient often communicates other information in addition to the location. However, the interviewer only asks about the location. The interviewer then recapitulates: So you're having pain in your knee, neck, shoulder and back? Do you have pain in other places? Each complaint is subsequently attended to: Is it continuously present? In what circumstances is it present? What is the exact location? etc.

In the case of psychological problems, it is important to inquire about psychotic symptoms (hallucinations and delusions) because the patient often does not mention these spontaneously.

General anamnesis and medical history:

In addition to the primary complaint, there may be other illnesses and restrictions that are important when assessing restrictions. It is often useful to send patients an anamnesis list in advance where they can note the complaints they have had and currently have, as well as past operations or hospital admissions. Clients can then take their time at home to list all the points, checking them if necessary.

why Clients often do not spontaneously mention complaints and restrictions that are not related to the primary complaint, especially if these have been around for a long time and have somehow become part of who they are.

example A patient comes for an assessment because psychological complaints have forced her to stop work, but she is also a patient with a chronic back condition. She doesn't spontaneously mention the back complaints because 'they have nothing to do with my burnout'. Besides limitations for mentally taxing work there may be limitations for work which is physically taxing.

Course of complaints:

Since when has the patient suffered from these complaints? How have they developed over time (especially in the months before the assessment)?

why The course of the complaints provides indicators for the prognosis.[16]

example The prognosis is not immediately favourable with regard to work that strains the back for someone who has suffered back complaints for 5 years and who has shown no improvement in recent months.

Cause:

Although it is the physician who often determines the cause of the illness, it is also important to know what patients themselves believe to be the cause. The cause is important for finding out whether the patient has received the right treatment, which restrictions are present and to test for consistency; it is also essential for assessing the prognosis.

Cause of psychological complaints:

For psychological complaints, the most complicated and elaborate diagnoses are often devised. Different physicians frequently fail to agree on a diagnosis. Fortunately, an insurance physician does not need to know the precise diagnosis to determine a patient's restrictions. Nevertheless, it may be useful to have an impression of the mechanism involved. A quick and informative method is to make a PPW (Private, Personal and Work) inventory. The idea is to list:

Private stressors: both negative and positive life events such as moving house,

cohabiting or getting married, relationship problems, having children, the death of friends or family, etc.

Personal factors: such as youth problems or parenting problems, neurotic problems, problems on the intrapyschological or character plane (e.g. biting off more than you can chew, perfectionism).

Work stressors: conflict at work, change in work content, high workload, restructuring, emotional demands, new job, etc.

why Understanding the mechanism of psychological decompensation can say something about the patient's psychological ability to cope with work and about consistency, as well as about the prognosis.[16]

example A patient suffers from burnout because his work has become increasingly pressurized following a restructuring (work). He is someone who finds it hard to say no and who readily bites off more than he can chew (personal). The patient mentally decompensates when, on top of that, his mother falls ill (private).

Treatment

Medication:

It is essential to be aware of all medications being taken.

why This indicates which illnesses play a role and their severity. The patient may also experience restrictions as a result of the medication.

example The pain is evidently serious if the patient takes morphine. If the patient takes antipsychotic drugs, this is not simply a question of burnout. There are restrictions on driving a car with medications that cause drowsiness.

Therapy:

What therapy has the patient undergone? What therapy is the patient still undergoing/ about to undergo?

why At the moment that the patient receives therapy, he or she cannot work. With psychological therapy, it is important to inquire about the emotional burden and the patient's response, as this can have implications for their ability to cope with work. If there have been many kinds of therapy and no sign of improvement, this may have implications for the prognosis. The nature and frequency of the therapy may also say something about the nature and severity of the illness.

example Any one undergoing five full days of treatment cannot work. With fewer days of treatment too, resuming work alongside the therapy is often contraindicated because the patient needs the remaining time to process the therapy.

For therapy that is emotionally demanding, the patient may not be able to work the

following day. It may be necessary to specify how many hours a patient can work to leave enough time to process psychological issues.

Opinion of physician:

What does the physician say to the patient about the diagnosis and the patient's ability to work?

why Coordination between the insurance physician and physicians gives the patient clarity. The physician may have arguments for a different view of the patient's capacity for work.

The diagnosis can say something about the prognosis.

example The physician may find someone fully unfit for work or permanently unfit for work that places strain in the back. It is important to know whether this is based on valid arguments.

The prognosis for back metastasis is different from arthrosis.

Examination

Physical and psychological examination:

In addition to regular physical and psychological examinations, the assessor also needs to pay attention to the patient's illness behaviour, for example How does the patient get up from the chair? How does she get dressed? Are the findings consistent? What is her walking pace and pattern? How does she sit on the chair? Does she give a firm handshake?

Psychological: alongside regular psychiatric examinations: Does the patient give adequate answers? Can he stick to the main focus of the interview? Is he emotionally stable? What does the assessor feel when speaking with the patient?

Information from physicians:

Information can be requested from treating physicians to supplement examinations that the assessors have conducted themselves. This information can help to assess 'external consistency'. It will not always help to identify a patient's restrictions because such information tends to have a curative focus, such as diagnosis and the benefit of treatment. A diagnosis is not strictly necessary for an insurance physician's assessment. Also, the patients themselves are often fully aware of the physicians' findings. See also the standard 'Communicatie met behandelaars' [Communication with physicians].[17]

Indicators for requesting information include:

- if it is not clear what is medically wrong
- if the patient requests it

- if the patient indicates that there is a difference of opinion between the treating physician and insurance physician
- if there are inconsistencies in the assessment
- if there is little understanding of the illness, simulation or dissimulation.

Further testing:
There are various possibilities for further tests such as Functional Capacity Evaluation (FCE), psychological testing, heteroanamnesis, questionnaires and an examination by a labour expert. See also the standard 'Onderzoeksmethoden' [Examination methods].[18]

Activity limitations experienced

Once an inventory of complaints has been drawn up, the assessor should inquire about any activity limitations that the patient experiences, especially in relation to work. Here too, in the case of multiple complaints, it is best to inquire in a structured way about the limitations the patient experiences. Questions about specific, detailed examples often give a good indication of severity (Appendix 1).[19] After all, patients often state in these examples the limits of what they can still do. Where possible, it is important for the patients to think up the examples by themselves because this indicates their severity and the impact on their lives. If the patient has problems finding everyday examples, then it is doubtful that he or she experiences serious restrictions.

why All doctors have their own specialisms and associated special anamnesis. A cardiologist will ask about complaints in the area of cardiology and a neurologist in the area of neurology. Insurance physicians need to identify restrictions in relation to work. Their special anamnesis includes questions about the patient's restrictions and activities.

example With regard to concentration problems, one patient says that she has problems with a difficult course of study or reading policy documents, whereas another patient finds it hard to read a simple magazine or watch a TV soap.
With regard to back complaints, one patient says that she cannot lift a briefcase and another has problems lifting paving stones.

Because patients do not spontaneously mention all restrictions, it is important to inquire about those that often accompany that syndrome. If necessary, the assessor can take a standardised work restrictions lis such as the Dutch 'Functional Ability List' (FAL)[20] and go over the restrictions point by point. This applies to physical and psychological complaints. A way to enquire about psychological restrictions without omitting any restrictions is to ask about the items on the Mental Ability List (MAL).[21,22] See also Appendices 2 and 3.

example For back complaints, patients often spontaneously mention problems with bending and lifting. Sitting, standing and walking are also regularly impaired and should be addressed in the anamnesis.
For psychological problems, patients often mention the complaints (sleep disorders, mood problems). However, the restrictions are important. Although not spontaneously mentioned, many psychological restrictions are often present, for example problems maintaining structure, not coping with emotional burdens or conflict, and concentration problems.

Participation

The goal is to gain an idea of the activities which patients still undertake despite the complaints and those which they no longer undertake. If patients no longer do things themselves, who helps and how often? The following activities can be enquired about:

- ADL, self-care
- managing their own administration and finances
- caring for and raising children
- housework, cooking, shopping, jobs in and around the house
- hobbies and leisure time activities
- sport
- work
- social contacts and functioning with the (extended) family, and others
- account of a day

Here too it is important to continue asking questions about specific and detailed activities (Appendix 1). In the section on activity limitations experienced by the patient, the patient could indicate what he or she can no longer do as a result of the illness. This section shows what the patient is still capable of doing. It also reveals what the patient can no longer do.

Account of a day:
This is the patients' account of how they spend their days. The aim is to obtain as detailed as possible an impression of a day in the life of patients: When do they go to bed and get up? Once they are up, what do they do from one hour to the next? Do they lie in bed or on the sofa during the day, and for how long? A specific, hour-by-hour anamnesis of 'yesterday' [23] provides a lot of information. However, it is very intensive and you run the risk of coming across as a detective. Nevertheless, asking the patient to describe an average day will result in a more subjective account. The intensity of questioning will depend on how clear the restrictions are.

why An account of a day tends to place greater emphasis on things that patients can still do. This can to some extent act as a counterbalance to only asking about complaints and restrictions. For the rest, it is a powerful tool for checking the consistency of a patient's story, and it often gives a good picture of the severity of the restrictions.

example A patient with Multiple Sclerosis performs light housework in the morning and has to go to bed in the afternoon in order to have the energy to undertake some activities in the evening. This could indicate a restriction in the number of hours that the patient can function per day. A patient with psychological complaints sees herself as being able to work half days only but she can fill an entire day doing housework and caring for the children. There appears to be a case for regarding her as capable of working full days.

Patient's view

Here patients can say what they themselves consider their restrictions and possibilities to be. It needs to be clear, for instance, why patients believe that they cannot do their job, and what they think of their options for other, perhaps lighter, work. The physician must make it clear that the assessment refers to restrictions for work in general (not just specifically the patient's own work) and that a declaration of incapacity for work takes no account of capacities in the home situation. The physicians can also offer here a preliminary impression of the patient's capacity for work.

why This feedback can give rise to new restrictions or a nuancing of the restrictions. Clients may raise arguments as to why they do not agree with the physician's provisional opinion.

example A patient considers himself able to work half days because he can only do his own job for half a day. He has psychological complaints and finds it hard to concentrate in his work as a proofreader. The physician deems the patient capable of working a full day, provided the work does not involve intense concentration. Following an explanation from the physician, the patient also considers himself capable of working a full day as a bridgeman, doorkeeper or similar occupation requiring no intense concentration.

Physician's opinion

On the basis of the available data, the physician decides whether additional information is needed and assesses the patient's capacity for work in accordance with rules and regulations, standards and protocols. This assessment is not a matter of recording what patients say that they can do, but what the physician believes they are still able to do. Important considerations include consistency and plausibility. The assessment

is communicated to patients clearly and in language that they understand. If it is not yet possible to make an assessment, agreement is reached on how and when this will happen.

Example There is no consistency between disorders, restrictions and handicaps if the patient indicates that he feels a lot of pain in his shoulder and can't lift anything yet continues to do motocross. There is a lack of consistency with regard to restrictions if the patient claims to be able to pick up a full bucket of water but not a packet of sugar. Here are two examples of reaching a decision on the items 'sitting' and 'concentrating'.

Sitting

Client says:	I can sit for no more than 15 minutes because of back trouble.
Restriction:	Back pain when sitting in church and at birthday parties.
Participation:	Drives from Groningen to Amsterdam (200 km) with a stop in Zwolle (at 100 km). Goes to church and to the cinema.
Examination:	Sits on the chair for ½ hour with no support and without shifting position.
Diagnosis:	Chronic aspecific back complaints.

Insurance physician's conclusion: the patient can sit for one hour

Reasoning: The patient can sit for 1 hour in a fixed position in the car. Although this is not for an entire day, the patient does not avoid sitting for longer periods, for example in church or at the cinema. The examination did not show that the patient can only sit for a shorter period. Given the diagnosis, this should be possible.

Concentrating

Client says:	I can't concentrate at all because I'm burnt out.
Restriction:	Reading magazine is fine, but can only manage the simpler summaries and headlines, not the longer items in a newspaper. Can read two pages of a detective novel in one sitting.
Activities:	Reads a magazine for a maximum of ½ hour. Watches television for a couple of hours, does watch *The Bold and the Beautiful* but no longer watches *Crime Scene Investigation*.
Examination:	Can focus well during a ½-hour interview. Is able to understand reasonably well the explanation of the assessment regarding work capacity, which is fairly complex.
Diagnosis:	Burn out

Insurance physician's conclusion: can cope with work involving short-term, specific attention and concentration on a particular task.

Reasoning: Client can follow simple texts and TV programmes, but not lengthy or complicated ones. This matches the diagnosis of burn out.

Specific questions for the patient:

In general, the questions should be tailored to the level of the patients. Clients will find some questions difficult to answer because of psychological barriers. It may be necessary to ask several questions or to incorporate further questions in the interview. Not all questions will be immediately understood by patients and will require explanation in advance (e.g. the question about the patients' own opinion of their ability to cope with their customary work). Interview skills are needed.

Work:

- What does your job entail?
- How long have you been doing this work? Has much changed?
- What are the difficult parts of your job? (how would you explain this to a friend who would like to do the job?). Both physical and psychological.
- What was the atmosphere like at work (with colleagues and management)?
- Were there problems or friction at work?
- Do/Did you enjoy the work?

Illness

History of illness:

- When did the complaints begin?
- Can you describe what else has happened up until the present?

Cause:

- What is the GP's or specialist's diagnosis?
- For psychological complaints: Did something happen at home? Are these things still happening? How would you describe yourself?

Course of complaints:

- How long have you been experiencing problems? How long did you have these problems before you stopped work?
- Has there been any improvement in the past few months?

Nature:

- What problem are you experiencing?
- Are you otherwise in good health?
- Have you been operated on in the past?
- Are you still experiencing any problems as a result of the operation?

Treatment:

- Are you still under treatment? If so, what treatment and how often?
- What form did the treatment take?
- Are you still expecting treatment?
- What medications do you use?
- Are you still seeing the specialist or your GP?
- When was your last appointment?
- Did the specialist/GP offer any advice?

Examination:

How does the patient get up from the chair (in the waiting room, going to the examination room and upon leaving)?

Does the patients sit still on the chair? For how long? Or does he/she keep changing position or gets up?

What verbal and non-verbal clues does the patient give during the functional examination?

How does the patient picks up his bag? How does he/she shake hands?

Further physical examinations targeting the syndrome.

Can patients answer the questions satisfactorily? Do they have to think for a long time? Have they forgotten a lot? Does the patient come alone? What is the interaction like with the patient's partner?

General psychiatric examination.

Activity restrictions:

- What things can you no longer do because of your illness?
- You have problems with …….(complaint). What does that mean on a day-to-day basis?
- What can you do? What can't you do?
- Can you give an example? (a concrete example for each restriction, persevere with questioning).
- How often does it occur? (ask for a specific answer: How often on average per week? How long on average does it last?)
- If it occurs, what do you do? (specific answer needed. Not: *I don't do anything*, but: *I lie in bed sleeping/I lie in bed reading*)

- Do you also have problems with ……? (the insurance physician fills in those restrictions that he or she believes accompany the ones already mentioned, e.g. sitting for those with back problems. If necessary, go through the FML and Mental Ability List (Appendix 2) point by point)
- Do certain activities make the complaints worse?

Participation:

- What are your hobbies and are there ones you no longer do?
- Do you still play sport?
- What housework do you do? (Did you do more in the past?)
- Do you still work? If so, what do you still do?
- What jobs do you do around the house and what do you no longer do?
- Are you married/cohabiting and do you have children (if so, how many and how old are they)?
- How are the relationships with your family, friends and acquaintances?
- Do you have help for certain things that you can no longer manage alone? If so, from whom and how often?
- What is your day like?
    - At what time do you get up and go to bed?
    - Do you lie in bed or on the sofa during the day?
    - What do you do in the morning? In the afternoon? In the evening? (be specific)

Client's view:

- What aspects of your work can you not manage at present?
- If you were to do your job, what would go wrong?
- Do you think you could work at a different job? If so, what conditions would have to be met?
- Listening to your story, I think that you would be able to …..…….(physician's provisional opinion of patient's work capacity). Do you agree? If not, why not?

Physician's opinion:

- If I list everything, I find that you have restrictions for ……(identify the key restrictions in the work pattern).
- This means that you are fit for ……(summarize the possibilities, e.g. work that places little strain on the back).
- This means that ……(possible implications for benefit)
- The next step is …….. …….(e.g. contact with an labor expert)

Reference List

1. International classification of functioning, disability and health. Geneva: World Health Organisation, 2001.
2. Boer WEL de, Brenninkmeijer V, Zuidam W. Long-term disability arrangements. A comparative study of assessment and quality control. Hoofddorp: TNO work and employment, 2004.
3. 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;25:501-12.
4. Boot CRL. Sick leave in astma en COPD: the role of the disease, adaptation, work, psychosocial factors and knowledge. Nijmegen: Dissertation, 2004.
5. Hodiamont PPG, Hodiamont-Kroon AME. Besliskunde in de forensische psychiatrie. [Decisionmaking in forensic psychiatry].Tijdschrift voor psychiatrie 1997;39: 484-9.
6. Boer WEL de, Wijers JHL, Spanjer J, Beijl I van der, Zuidam W, Venema A. Gespreksmodellen in de verzekeringsgeneeskunde. [Interview models in insurance medicine]. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2006;14:17-23.
7. Schattingsbesluit arbeidsongeschiktheidswetten. [Assessment Decree Occupational Disability Act]. The Hague Netherlands, Ministry of Social Affairs and Employment, 2004.
8. Köhnken G. Statement validity analysis and the detection of the truth. The detection of deception in forensic contexts pp 41-63. New York: Granhag, Pär-Anders, Strömwall, Leif, 2004.
9. 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 and criminological psychology 2005;10:225-45.
10. Vrij A. Criteria-based content analysis: a qualitative review of the first 37 studies. Psychology, Public Policy and Law 2005;11:3-41.
11. Tredouw C, Foster D, Allan A, Cohen A, Wassenaar D. Psychology and Law. Lansdowne, South Africa: Juta and Co Ltd, 2005. Chapter 6 Tredoux C, Pooley S. Credibility Assessment.
12. Lamb ME, Hershkowitz I, Sternberg KI, Esplin PW, Hovav M, Manor T, Yudilevitz L. Effects of investigative utterance types on Israeli children's responses. Int J Behav Dev 1996;19:627-37.
13. Spanjer J. De betrouwbaarheid van WAO beoordelingen. [The inter- and intra- rater reliability of disability pension assessments]. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2001; 9:234-241.
14. Spanjer J, Krol B, Brouwer S, Groothoff JWG. Inter-rater reliability in disability assessment based on a structured interview report. Disability and Rehabilitation 2008;30:1885-90.
15. Is beoordeling op basis van een schriftelijke rapportage betrouwbaar? [Inter-rater reliability based on written reports]. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2005:13;105
16. Spanjer J. Professionele prognosestelling verkend. [Prognosis of functional limitations].Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 1998;2:41-6.
17. Communicatie met behandelaars (standaard). [Communication with treating physicians (protocol)]. Naslag gevalsbehandeling sociale verzekering. Houten/Diegem: Bohn Stafleu Van Loghum, 1996.
18. Spanjer J. Voorbeelden: een hulpmiddel om de ernst van de beperkingen te beoordelen. [Examples: a way to assess the seriousness of work limitations].Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2002;10:109-111.
19. Functionele mogelijkheden lijst. [Functional Ability List]. Amsterdam: UWV, 2002.
20. Spanjer J. De praktische concretisering van psychische belastbaarheid. [Documentation of mental abilities]. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 1996;5:182-6.
21. De speciële anamnese bij psychische beperkingen. [Anamnesis into mental functional limitations]. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2005:13;277-8.
22. Lomax CL, Brown RG, Howard. Measuring disability in patients with neurodegenerative disease using the 'Yesterday Interview'. Int J Geriatr Psychiatry 2004;19:1058-64.

BACKGROUND

After working as an insurance physician for about six years, I began providing practical supervision to physicians who had just started work. Those were turbulent times with many restructurings, and there were always many physicians needing training (on average about four per year). I advised them on how to conduct an assessment interview. In my discussions with these insurance physicians in training, it became increasingly clear to me what I myself did in an assessment interview and why. After a while, I noticed that I was always having to explain the same things, which is why I have committed them to paper. This is how the Disability Assessment Structured Interview (DASI) was born. As a result of discussions with colleagues, I was able to make various sections of DASI more specific and to add additional sections. Other changes were made following discussions with insurance physicians who had almost completed their training and to whom I taught the assessment of psychological restrictions at the Netherlans School of Public Health (NSPOH). Further adjustments were made following the analysis of 14 video recordings of DASI assessment interviews that were made as part of a research study. When these video recordings were shown to 22 insurance physicians from various institutions, they were accompanied by a detailed commentary, part of which has been incorporated into DASI. In the meantime, the practical feasibility of the method was constantly being tested through practical application.

In January 2001, DASI was selected as the preferred method in assessment interviews conducted by 'non-physicians' at Cadans. These staff members received three days' training in DASI, while their supervising physicians were given one day. Because the physicians supervised their colleagues and had to make an assessment on the basis of their reports, it became apparent that the Examples section in particular needed elaboration. In addition, because I had to describe DASI to non-physicians, I was obliged to explain things that a physician tends to do automatically. This made it clear to me that physicians have automatisms that may interfere with an objective assessment.

I conducted a literature review of standardized interviews to ensure that DASI had a sound research basis, and to further tighten up some sections (e.g. the legal literature for statement credibility).

Appendix 1

Examples: a way of determining the severity of the restrictions

A patient with considerable back pain has been diagnosed by the orthopaedist as having lumbar arthrosis and can no longer work as a construction worker. The patient enters, walking with difficulty. He says that the orthopaedist has advised him to stop work. He is not allowed to lift heavy objects. He is given weekly physiotherapy (paraffin packs and massage). A physical examination of his back reveals the following:
Posture: light scoliosis, straight pelvis
Function: flexion, extension and rotations of the back are possible but difficult
No indications of radicular irritation.

The insurance physician is asked to assess how much the patient can lift and how often. Is there enough information at present to answer this question? In my view, there isn't. Although it is plausible that there are strong restrictions on lifting, their severity cannot be accurately determined. The syndrome or the severity of the complaints does not say much about the severity of the restrictions. How do you establish the severity of the restrictions that someone is experiencing?

After the patient is asked which restrictions he has, the anamnesis can focus on their severity. A good way to do this is to ask for examples of the restriction and of the activities that the patient still does or does not undertake. With such examples, patients indicate the limits of what they can do. Clients do not give examples of things that they can do very easily or only with great difficulty, but of things that they can only just manage or only just fail to do. This limit is valuable information, which the insurance physician should record in the assessment and the work pattern.

The patient says that he finds it hard to lift things and that three kilos is about the maximum. The patient gives several examples of his lifting problem. He can only pick up his grandson once in a while. His grandson is one year of age and by 'once in a while' he means at most three times a day. He then feels more pain in his back, which stays with him for the rest of the day; sometimes he even has to go to bed. Another example is that his wife has to lift the weekly crate of beer. Even with two hands he can't do it any longer as it puts his back out. If he works in the garden, he can no longer pick up a full bucket of water. Instead, he lifts two half-full buckets, one on each side. He can't do this more than about four times a day, and he uses the garden hose more often than he used to.

Several things are important when asking for examples. I will discuss a few of them below, and explain them on the basis of the above example.

Specific: The insurance physician needs to ask for concrete examples. Instead of saying: I lift very little, and occasionally a little bit more, the patient should be encouraged to say: I pick up my grandson, I lift 2 half-full buckets of water and I don't lift a crate of beer.'

Try to make vague utterances as specific as possible by asking: What is occasional lifting? What is sometimes? What is for a short time? What is regularly? Specific questions should be asked: how often? how much? how long? Otherwise there could be differences in interpretation between the insurance physician and the patient. In the above example, the physician asks what the patient understands by occasionally (3 times a day).

The patient's estimates are not always correct: standing for 5 minutes, lifting 3 kgs and walking 1 km are rough estimates. Based on the examples the patient mentions, the physician can make a more objective estimate, which doesn't necessarily correspond to the patient's assessment of the severity of the restriction. In the above example, the patient estimates that he can lift 3 kgs, while the physician, on the basis of the examples, puts this at 10 kgs. This is the difference between assessing and recording what the patient says.

The clearer and more plausible the restriction, the fewer additional questions need to be asked. The less clear the restriction (if it doesn't match the complaints or the diagnosis, if there is a discrepancy between subjective and objective matters), the more examples need to be elicited and the more specific the questions should be.

Detailed: The more details there are, the more accurate the assessment of the restriction can be. How old or how heavy is your grandson? How big is the crate of beer (24 bottles)? Is it empty or full? What is in the buckets? The clearer and more plausible the restriction, the less detailed the questions need to be.

Not every component of the work pattern has to be inquired about. Sometimes the insurance physician can reflect on the presence/absence and severity of the restriction.

Spontaneous: It is best if the patients themselves come up with examples. You can assume that people with restrictions will be confronted with them in daily life. Clients should therefore mention these examples spontaneously. If the restrictions are not serious, or are even simulated, this will be more difficult. It is true, however, that the longer patients suffer from restrictions, the more they integrate them into their lives and no longer view them as such. Someone with a social phobia, whose restriction is that she has problems being among other people and who for years has adapted her life accordingly, will not always spontaneously mention this restriction as it has become part of everyday life.

A patient will not always spontaneously give a full description of all restrictions. The physician will have to enquire about restrictions that are not mentioned. Key items in the work pattern can be inquired about. For example, patients with psychological complaints may spontaneously say that they cannot cope with responsibility and time pressure. On further questioning, however, it may emerge that there also problems with concentration and dealing with conflict.

Ideal: the patients themselves come up with examples of restrictions. Give examples of things you have difficulty with.
Good: the physician comes up with the restriction, and the patient gives examples. And what's your concentration like? Give some examples.
Not so good: the physician comes up with the restriction and guides the examples. How do you find reading or watching TV?
Bad: the physician comes up with the examples. No doubt you experience problems reading and watching TV for a long time.
Recent: Inquire about recent examples of restrictions. After all, we are interested in what they are like now. Especially if things are starting to improve, the patient may tend to say how bad it used to be.
Consistency: The examples need to show consistency with other examples and of course with other matters, for example the physical examination and the syndrome. The above example shows consistency in terms of the weight that the patient is still able to carry: a one-year-old grandson weighs approx.10 kgs, which is the same as two half-full buckets of water. A crate of beer is about 15 kgs, which is too heavy. The frequency with which this maximum of 10 kgs can be lifted is also consistent: the grandson is picked up 3 times a day at most, and the bucket about four times a day.
Why: It may be useful to know what happens if the patient exceeds this restriction. This may help to reveal whether the behaviour is based on illness. Compare, for instance, I don't vacuum because I hate it with I don't vacuum because my back starts hurting. It can also be useful in relation to therapy advice. Compare, for instance, I don't lift more than two kilos because I'm afraid that my hand will hurt with I don't lift more than two kilos because otherwise my hand hurts. In the above example, what would happen if the patient did pick up a crate of beer?

Appendix 2. Mental Ability List

Descriptions of the assessment points for a patient's psychological ability to cope with work
(For each of the eight items, circle the level that represents the maximum that the patient is capable of)

1. STRUCTURE

1. Should do highly structured work, with fixed tasks and a fixed work schedule, and with no unexpected duties or other matters. No need or opportunity to impose own structure.
2. Is capable of work involving the occasional small duty in between, but for the rest structured work with a reasonably fixed time and task schedule. No need to impose own structure.
3. Is capable of doing work where a proper structure can be imposed (and which is to some degree necessary to avoid chaos), and involving some unexpected duties.
4. Is capable of work that regular involves unexpected matters, only a small part of which can be structured.
5. Is capable of hectic work involving many unexpected matters (phone calls, requests). It is impossible to impose a structure.

2. RESPONSIBILITY

1. Should do work involving no responsibility. There is always someone to check if things go wrong. Nothing can go wrong.
2. Is capable of work with little responsibility. Not much can go wrong, or there are regular checks.
3. Is capable of working with moderate responsibility. No ultimate responsibility. Mistakes do not have major consequences.
4. Is capable of working with a high degree of responsibility and must take important decisions with major consequences. However, there is either considerable opportunity for consultation or a superior has ultimate responsibility.
5. Is capable of working with a very high degree of responsibility, has ultimate responsibility in important decisions with major consequences.

3. TIME PRESSURE

1. Should do work with very little pressure. The tasks are easy to perform at a calm pace.
2. Is capable of working under moderate pressure, with sufficient breaks and no production peaks.
3. Is capable of working under average pressure, with a normal working pace and opportunities for relaxation.
4. Is capable of working under high pressure, can continue working at a considerable pace with little or no time to relax.
5. Is capable of working under very high pressure, with very tight schedules, and at a constantly high pace.

4. EMOTIONAL BURDEN

1. Should not deal with the problems of others.
2. Can cope with the problems of others, provided they are not serious.
3. Can cope with major suffering (e.g. chronic illness and death) of others.

5. CONCENTRATION

1. Can maintain attention during simple, routine work involving no specific demands on concentration.
2. Can cope with work requiring specific, short-term attention to and concentration on a certain task.
3. Can concentrate for long periods, provided the subject matter is not complex.
4. Can concentrate intensely for long periods on complex subject matter.

6. ENVIRONMENT

1. Needs a work environment that is calm, quiet and transparent.
2. Is capable of working in a rather quiet environment, with occasional noises from someone coming and going.
3. Is capable of working in an environment with regular noises or someone coming and going.
4. Is capable of working in a fairly restless environment with a lot of noise.
5. Is capable of working in a hectic, noisy and chaotic environment.

7. CONFLICT

1 Cannot cope at all with conflict.
2. Can sometimes cope with a minor conflict.
3. Can cope with conflict provided this is not face-to-face (e.g. by phone or on paper), or with face-to-face conflict that is not too intrusive.
4. Can cope with face-to-face conflict but not with aggressive or unreasonable people.
5. Can cope with face-to-face conflict and with unreasonable or aggressive people.

8. RELATING TO OTHERS

1. Should work more or less alone.
2. Can work with a small group of familiar people, not with strangers.
3. Can work with a fairly stable group of people, and occasionally with strangers.
4. Can work regularly with strangers, occasionally with a more stable group.
5. Can work in large groups with strangers or constantly changing people.

9. OTHER RESTRICTIONS

Restrictions that do not come under the above items.

_____

_____

_____

Restrictions in work pattern:

Hours per day: _____

Days per week: _____

Other: _____

Date: _____

Physician: _____

Appendix 3. Anamnesis with psychological restrictions

Structure

- Can the patient make plans (e.g. in relation to housework, shopping or work-related matters)?
- Can the patient adhere to those plans?
- What happens if the structure is disrupted or unexpected events arise?
- Does the patient have difficulty doing several things at once (e.g. cooking dinner and seeing to the children)?
- Are third parties needed to impose structure?

Responsibility

- Does the patient make decisions independently, and if yes, which ones? e.g. about housework, children, work, etc.
- Can the patient also make an independent decision if that decision has major consequences?
- Which decisions are difficult?
- Are checks or consultation needed for certain decisions?

Time pressure

- Is it a problem if something has to be done within a certain time?
- Does the patient need to have breaks?
- Can the patient manage if something has to be done more quickly (rapid pace)?
- Can the patient cope with several appointments in one day?

Emotional burden

- Can the patient bear to listen to or observe the problems of others?
- Can the patient distance him or herself from this?

Concentration

- Can the patient focus attention on something and maintain attention on certain things?
- Can the patient do this for a longer period?
- Can the patient do this if the subject matter is complicated?

Environment

- Can the patient cope with noise and commotion in his or her vicinity?
- Are there problems with loud or much noise?
- Are there problems if there are many people in his or her vicinity?

Conflicts

- Can the patient deal with conflict and opposition?
- Can the patient stick up for him or herself or defend his or her point of view?
- How does the patient manage this in writing, over the phone, or face-to-face?
- How does the patient manage this with 'difficult' (i.e. aggressive or unreasonable) people?

Dealing with others

- Can the patient get on well with other people?
- Can the patient get on well if there are strangers present or as part of a group?
- Are there problems with certain people?

Addendum 3

Belastbaarheid Gericht Beoordelingsgesprek (BGB)

Versie 5.2 (Oktober 2007)

Jerry Spanjer

SAMENVATTING

De Belastbaarheidgerichte beoordeling (BGB) is een semi-gestructureerd interview om gegevens te verzamelen voor de beoordeling van de belastbaarheid voor arbeid en de prognose daarvan. Bij een beoordeling zijn naast dit interview andere gegevens van belang zoals observatie, onderzoek en de informatie van de behandelaars. In het interview worden gestructureerd de ICF[1] domeinen: stoornissen, beperkingen en participatie in kaart gebracht. Belangrijk onderdeel daarbij is de anamnese naar concrete en gedetailleerde voorbeelden van belemmeringen en nog ondernomen activiteiten. Op basis van de verzamelde gegevens kan de belastbaarheid voor werk worden beoordeeld.

BELASTBAARHEID GERICHT BEOORDELINGSGESPREK

Een semi-gestructureerd interview gericht op de beoordeling van de belastbaarheid voor werk

INLEIDING

In zowel Europa als in de Verenigde Staten wordt de beoordeling van de belastbaarheid verricht door medici, Nederland is in zoverre uniek dat het werkt met een geprotocolleerd interview[2]. Het beoordelen van de belastbaarheid is echter niet een onderdeel van de artsenopleiding. In de opleiding tot basisarts wordt geleerd te werken volgens een vast stramien: na de anamnese en het onderzoek volgen hulponderzoeken zoals laboratoriumonderzoek en röntgenfoto's. Op basis van die informatie wordt een diagnose gesteld en op basis van die diagnose kan een therapie worden ingesteld.

Verzekeringsartsen hebben niet als hoofdtaak de mensen te behandelen middels een therapie, maar de taak de belastbaarheid en de prognose van die belastbaarheid van de cliënt te beoordelen. Dit vereist een andere methodiek dan geleerd wordt tijdens de opleiding tot basisarts omdat de diagnose niet rechtstreeks gerelateerd is aan de aard en de ernst van de beperkingen. Zo staan bij de diagnose reumatoïde artritis de aard en de ernst van de beperkingen van een patiënt niet vast[3] en is aangetoond dat bij COPD de parameters die de ernst van de ziekte aangeven zoals de mate van longobstructie (FEV1) niet is geassocieerd met de ernst van de beperkingen.[4] Een ander verschil tussen curatieve artsen en verzekeringsartsen is het feit dat de curatieve arts te maken heeft met patiënten die bij een juiste vaststelling van de diagnose is gebaat en de verzekeringsarts te maken heeft met cliënten die niet altijd bij een juiste beoordeling van de belastbaarheid is gebaat. I.v.m. financiële of arbeidsrechtelijke consequenties kan een cliënt bijvoorbeeld zaken minder rooskleurig voordoen dan ze zijn.

Omdat het oordeel van de beoordelende verzekeringsartsen voor een groot deel wordt gebaseerd op het gesprek met de cliënt en gestandaardiseerde informatieverzameling van belang is voor

een lage inter-dokter variatie[5] is een semi-gestructureerd interview van belang. Er zijn momenteel een drietal min of meer geprotocolleerde gespreksmodellen beschreven waarvan hier 'het Belastbaarheid Gericht Beoordelingsgesprek' (BGB) wordt besproken.[6]

De BGB is een interview model om semi-gestructureerd gegevens te verzamelen met betrekking tot beperkingen bij ziekte en mogelijkheden in arbeid. De gegevens die de arts verzamelt kunnen divers zijn zoals: mondelinge informatie van de cliënt, lichamelijk en psychisch onderzoek, informatie van de behandelaar of werkgever en hetero-anamnese. Daarnaast kunnen diverse hulponderzoeken worden verricht zoals: expertise bij een specialist, psychologische testen, performance testen (ook wel functional capacity evaluation of FCE genoemd) en het afnemen van vragenlijsten. De nadruk bij de BGB ligt op het semi-gestructureerde interview van de cliënt.
Op basis van de totale verzamelde gegevens kan een beargumenteerd oordeel over de belastbaarheid in arbeid worden gegeven. De BGB beschrijft slechts welke gegevens verzameld dienen te worden en op welke manier, niet hoe het oordeel zelf geveld dient te worden. Hoe die beoordeling dient plaats te vinden staat beschreven in onder andere het 'Schattingsbesluit arbeidsongeschiktheidswetten[7], verzekeringsgeneeskundige richtlijnen, standaarden en protocollen. Verder zijn er invloeden van bijvoorbeeld de stand van de wetenschap (bijvoorbeeld: 'bij rugklachten is bewegen goed ipv bedrust') en intercollegiale toetsing. Desondanks blijven er discussiepunten bijvoorbeeld hoe om te gaan met moeilijk objectiveerbare klachten. Wellicht dat nieuwe instrumenten, zoals mediprudentie (voorbeelden van beoordeling, denk ook aan jurisprudentie bij rechterlijke uitspraken), nog een rol kunnen spelen om de variatie tussen beoordelaars te verkleinen. In elk geval blijft het van belang dat beoordelaars zich zoveel mogelijk baseren op dezelfde gegevens en is daarom een semi-gestructureerd beoordelingsgesprek van belang.

INHOUD EN UITGANGSPUNTEN

In het 'Schattingsbesluit arbeidsongeschiktheidswetten' wordt beschreven dat stoornissen (op het niveau van lichamelijk en geestelijk functioneren), beperkingen (op niveau van gedragingen en activiteiten) en handicaps (op niveau van sociale rolinvulling) een logisch en samenhangend geheel moeten vormen, er moet sprake van consistentie zijn. Dit sluit aan bij de domeinen van de ICF. In de BGB worden daarom die drie domeinen nadrukkelijk en systematisch uitgevraagd:

- Stoornissen: worden in kaart gebracht door gestructureerd naar de ziektegegevens te vragen en daarnaast het lichamelijk en psychisch onderzoek te verrichten.
- Beperkingen: gestructureerd wordt navraag gedaan naar concrete en gedetailleerde voorbeelden van ervaren belemmeringen.
- Participatie: gestructureerd wordt in kaart gebracht wat de cliënt wel en niet meer onderneemt in huishouden, sport en hobby's, ADL, sociale contacten en werk. Daarnaast wordt een gedetailleerd dagverhaal van de cliënt afgenomen.

In de juridische literatuur worden handvaten aangereikt om te beoordelen of een anamnese geloofwaardig is[8,9]. Een belangrijk onderdeel daarvan is een analyse van de inhoud van de anamnese: the Criteria-based content analysis (CBCA)[10,11]. In de CBCA worden criteria aangereikt om een anamnese op geloofwaardigheid te beoordelen. De CBCA gaat ervan uit dat als iemand iets zelf heeft ervaren dat de verklaring die hij aflegt verschilt van iemand waarbij dat niet het geval is. Alleen als iemand iets zelf heeft ervaren dan kan worden voldaan aan de CBCA karakteristieken. Algemene karakteristieken zijn: ongestructureerd het verhaal kunnen vertellen, de hoeveelheid details en een logische structuur. Kenmerken van de BGB zijn daarom:

- Gestructureerd uitvragen: Consistentie kan worden onderzocht door te beoordelen of op een aantal samenhangende domeinen (stoornissen, beperkingen en participatie) dezelfde aard en ernst van beperkingen aangegeven worden. Het is daarom ook van belang de anamnese onderdelen zoveel mogelijk in een vaste volgorde uit te vragen en los van elkaar. Door die structuur dient de cliënt zijn verhaal ongestructureerd te kunnen produceren. Als een cliënt bijvoorbeeld aangeeft weinig te kunnen tillen dan dient dat later in het gesprek terug te komen bij bijvoorbeeld de huishouding of het dagverhaal.
- Doorvragen naar concrete en gespecificeerde details: Door zoveel mogelijk te concretiseren en naar details te vragen is simuleren of dissimuleren moeilijker. Tevens kan de consistentie tussen de onderdelen beter onderzocht worden. Bovendien is te verwachten dat de variatie in het oordeel tussen de beoordelaars af zal nemen omdat tussen de beoordelaars minder interpretatieruimte zal bestaan. Als de cliënt bijvoorbeeld aangeeft 'regelmatig' hoofdpijn te hebben is meer interpretatieruimte mogelijk dan dat de cliënt zegt '1 keer per maand' hoofdpijn te hebben. In bijlage 1 wordt dit verder toegelicht en uitgewerkt.

In een beoordelingsgesprek spelen de normen en waarden van de beoordelaar zelf een rol en daarnaast ook diverse psychologische mechanismen. Voorbeelden hiervan zijn het 'rank effect' dwz vorige beoordelingen beïnvloeden de beoordelingen daarop en 'confirmation bias' dwz de neiging om informatie te zoeken of te interpreteren zodat die aansluit bij een al gevormd oordeel en andere informatie juist te negeren. Ook hier kan structuur en doorvragen naar details zorgen voor minder beïnvloeding van de beoordelaar, verder is een ander kenmerk van de BGB daarom:

- Alle gespreksonderdelen dienen te worden uitgevraagd: Door alle onderdelen uit te vragen vult de beoordelaar niet zelf in wat hij al denkt te weten,er zullen daardoor minder vooroordelen van de beoordelaar in de anamnese sluipen.

Het is niet altijd makkelijk om de structuur van het gesprek vast te houden of door te vragen naar concrete voorbeelden. Dit wordt veroorzaakt door diverse factoren zoals het feit dat de cliënt soms andere zaken wil vertellen dan de beoordelaar wil horen of zich moeilijk aan de structuur kan houden. Daarnaast worden aan de beoordelaar eisen gesteld met betrekking tot gesprekstechnische vaardigheden en is verstoring van de structuur mogelijk omdat de neiging kan

bestaan om tijdens de anamnese het gedrag van de cliënt te willen beïnvloeden ('When you are an interviewer, you are not a therapist'). In elk geval zijn gesprekstechnische vaardigheden zoals omgaan met weerstanden, open vragen kunnen stellen en het kunnen doorvragen een belangrijk instrument bij de anamnese[12]. Oefening en ervaring zijn daarvoor noodzakelijk.

De BGB is een model om relevante gegevens voor de beoordeling van de belastbaarheid en de belastbaarheidprognose te verzamelen. Een ander punt is de beoordeling zelf. Tussen de beoordelaars kunnen verschillen bestaan in de interpretatie van de gegevens. Naast een gestructureerde anamnese zijn daarom andere instrumenten zoals standaarden en richtlijnen nodig. Uit onderzoek naar de inter-dokter variatie met gebruik van de BGB bleek een redelijk betrouwbaar oordeel mogelijk[13,14,15]. De validiteit is moeilijk te meten omdat er geen gouden standaard is. Omdat de structuur van de ICF wordt gevolgd en rechtstreeks en concreet naar beperkingen wordt gevraagd in de anamnese i.p.v. ze af te leiden uit een diagnose kan een redelijke validiteit verwacht worden.

BELASTBAARHEID GERICHT BEOORDELINGSGESPREK

A. Introductie

- Voorstellen
- Op gemak stellen
- Korte uitleg van de procedure
- Samenvatten welke gegevens al bekend zijn.

B. Werk

- Aard en duur van dienstverband
- Werkinhoud
- Werkbeleving

C. Stoornissen

1. Ziektegegevens:

- Ziektegeschiedenis en aard van de klachten
- Tractusanamnese, voorgeschiedenis
- Beloop
- Oorzaak (PPW, diagnose behandelaar)
- Behandeling (nu en in verleden, medicatie, visie behandelaar)

2. Onderzoek:

- Psychisch, lichamelijk
- Eventueel nog info behandelaar, heteroanamnese, expertise

D. Ervaren belemmeringen

- Aard
- Ernst (concrete gedetailleerde voorbeelden)

E. Participatie

- Wat doet de cliënt wel en wat niet (meer)? (dagverhaal, sport en hobby's, huishouden, werk, persoonlijk en sociaal functioneren)

F. Visie cliënt

- Wat vindt de cliënt zelf van zijn mogelijkheden om te functioneren?
- Wat is de reactie van de cliënt op de voorlopige visie van de verzekeringsarts?

G. Visie verzekeringsarts

- De verzekeringsarts legt uit wat zijn oordeel is en wat de mogelijke consequenties hiervan zijn.
- Uitleg van de verdere gang van zaken.
- Ruimte om te reageren en afronding.

Werk

Naast de inhoud van het werk en de aard van het dienstverband (part time of full time) is ook de duur van het dienstverband van belang (hoe lang werkt de cliënt al bij deze werkgever in dit werk?). Verder wordt gevraagd hoe het werk werd ervaren door de cliënt. Van belang is welke onderdelen als (te) zwaar werden bevonden en hoe de sfeer op het werk was.

waarom Hieruit kan o.a. naar voren komen of er motivatie is om (niet) naar het werk te gaan. Verder kan onderzocht worden of er consistentie bestaat tussen het niet meer kunnen verrichten van (onderdelen van) het werk en de beperkingen. Als de cliënt het werk heel kort heeft verricht kan onderzocht worden of hij bij aanvang wellicht al ongeschikt was voor het werk.

voorbeeld Cliënt had veel problemen met een collega en de leiding op het werk. Hij heeft afwisselend licht administratief werk en is uitgevallen met lichte rugklachten. Er lijkt hier eerder sprake te zijn van een motivatieprobleem dan een gezondheidsprobleem.

Ziektegegevens

Ziektegeschiedenis en de aard van de klachten:

Het in kaart brengen van de geschiedenis van de huidige ziekte, de medische anamnese. Bij een enkelvoudige klacht is de noodzaak om de "teugels strak te houden" vaak niet zo groot. Bij multiple klachten kan een onontwarbare kluwen van klachten, visie van behandelaars, mogelijke diagnoses en oorzaken en verrichte behandelingen ontstaan. Het bieden van structuur kan dan behulpzaam zijn, door eerst te inventariseren welke klachtengroepen er zijn, en welke klachten in elke groep weer te onderscheiden zijn. Daarna kan (ook bij multiple klachten) gestructureerd verder gegaan worden met de anamnese per klacht. Zo kan gestructureerd en in relatief kort tijdsbestek een inventarisatie gemaakt worden en worden geen klachten gemist.

voorbeeld Een cliënt met fibromyalgie heeft twee klachten groepen nl. moeheid en pijn. De pijn bestaat uit pijn in de knie, nek en schouders en in de rug. Eerst wordt de cliënt gevraagd: waar heeft u pijn? Cliënt geeft dan naast de lokalisatie vaak diverse andere informatie door. Dan wordt echter gevraagd alleen eerst de lokalisatie door te geven. Na afloop wordt samengevat: U heeft dus pijn in de schouder, nek, en knie: heeft u verder nog ergens pijn? Daarna wordt een inventarisatie per klacht gestructureerd uitgevraagd: is de cliënt altijd moe?, is het continu aanwezig? etc. Na de moeheid de kniepijn: waar doet het pijn, wanneer treedt de pijn op, wordt de knie dik? Etc.

In geval van psychische klachten is het van belang navraag te doen naar psychotische fenomenen (hallucinaties, wanen etc.) omdat die vaak niet door de cliënt spontaan genoemd worden.

Tractus anamnese en voorgeschiedenis:

Naast de primaire klacht kunnen andere ziektes en beperkingen meespelen die van belang zijn voor het vaststellen van de beperkingen. Eventueel kan aan de cliënt voor het gesprek een anamneselijst toegestuurd worden waarin hij kan aangeven welke klachten hij allemaal heeft en heeft gehad en daarnaast welke operaties en ziekenhuis opnames in het verleden geweest zijn. Cliënt kan dan thuis rustig deze zaken op een rijtje zetten en eventueel nakijken.

waarom Cliënt komt vaak niet spontaan met klachten en beperkingen die niet met de primaire klacht te maken hebben, met name als die al heel lang bestaan en een beetje bij iemand 'horen'.

voorbeeld Iemand wordt gezien voor een beoordeling i.v.m. uitval met psychische klachten, maar is ook chronisch rugpatiënt. De rugklachten worden niet spontaan genoemd want 'die staan toch los van de overspannenheid?'. Bij functieduiding mogen echter geen rugbelastende werkzaamheden geduid worden ook al zijn die psychisch niet belastend.

Beloop:

Sinds wanneer spelen de klachten? Hoe is het beloop van de klachten in de tijd? Met name ook in de maanden voor de beoordeling.

waarom Het beloop geeft aanwijzingen voor de prognose van de beperkingen.[16]

voorbeeld De prognose voor rugbelastend werk bij iemand die al 5 jaar rugklachten heeft en de laatste maanden ook geen verbetering laat zien is niet direct goed.

Oorzaak:

De oorzaak van de ziekte zal vaak de diagnose van een behandelaar zijn, maar ook is van belang wat de cliënt zelf denkt dat de oorzaak is. De oorzaak is van belang om te onderzoeken of de cliënt de juiste behandeling heeft gehad, om te onderzoeken welke beperkingen aanwezig zijn en consistentie te toetsen en is van groot belang voor het beoordelen van de prognose.

Oorzaak van psychische klachten:

Bij psychische klachten kunnen ingewikkelde en uitgebreide diagnoses gesteld zijn. Soms zijn verschillende behandelaars het niet eens over een diagnose en is er discussie over de diagnose. Een verzekeringsarts hoeft die diagnose ook niet exact te weten om de beperkingen vast te stellen. Wel is het vaak nuttig om toch een indruk te hebben van het mechanisme. Een snelle en informatieve methode is inventarisatie van de "PPW". Die afkorting staat voor privé, persoon en werk. De bedoeling is een inventarisatie te maken van:

Privé stressoren: zowel negatieve als positieve life-events zoals verhuizing, samenwonen of trouwen, relatieproblemen, kinderen krijgen, overlijden van vrienden of familie etc.

Persoonlijkheidsfactoren inclusief jeugd- of opvoedingsproblemen, neurotische problematiek, intrapsychische stoornissen en karakter (bv teveel hooi op de vork nemen of perfectionisme).

Werk stressoren: conflicten op het werk, verandering van werkinhoud, hoge werkdruk, reorganisatie, emotionele belasting, nieuwe baan etc.

waarom Inzicht in het mechanisme van psychische decompensatie kan iets zeggen over de psychische belastbaarheid, consistentie en daarnaast over de prognose[16].

voorbeeld Een cliënt wordt overspannen omdat het werk een steeds hogere werkdruk heeft gekregen na een reorganisatie (werk) en de cliënt iemand is die moeilijk nee zegt en snel teveel hooi op de vork neemt (persoon). Cliënt is gedecompenseerd toen de moeder ook nog ziek werd (privé).

Behandeling

Medicatie:

Alle gebruikte medicatie dient bekend te zijn.

waarom Dit geeft een indicatie van welke ziektes meespelen en van de ernst van de ziekte. Verder kan de cliënt t.g.v. die medicatie beperkingen hebben.

voorbeeld De pijn is blijkbaar erg als de cliënt morfine heeft. Er is niet sprake van een eenvoudige overspannenheid bij antipsychotica gebruik. Beperkingen voor autorijden bij sufmakers.

Therapie:

Welke therapie heeft de cliënt gehad en welke therapie heeft de cliënt nu nog of volgt binnenkort?

waarom Op het moment dat de cliënt therapie heeft kan hij niet werken. Bij psychische therapie is van belang te vragen naar de emotionele belasting en de reactie van de cliënt hierop. Dit kan gevolgen voor de belastbaarheid hebben. Als er al allerlei therapie is geweest en de reactie in de vorm van verbetering is uitgebleven dan heeft dit mogelijk gevolgen voor de prognose. Verder kan de aard en frequentie van de therapie iets zeggen over de aard en ernst van de ziekte.

voorbeeld Iemand die 5 dagen dagbehandeling heeft is op die dagen niet belastbaar. Ook bij minder dagen

dagbehandeling kan werkhervatting naast de therapie gecontraïndiceerd zijn omdat naast de therapie tijd nodig is voor verwerking of het verrichten van opdrachten in het kader van de therapie.
Bij emotioneel belastende therapie is iemand soms de dag na therapie niet belastbaar. Ook kan het geïndiceerd zijn een urenbeperking in de belastbaarheid aan te geven zodat er tijd over is voor verwerking van psychische problematiek.

Visie behandelaar:

Wat zegt de behandelaar tegen de cliënt over de diagnose en belastbaarheid?

waarom Afstemming tussen verzekeringsarts en behandelaars geeft duidelijkheid aan de cliënt. De behandelaar kan argumenten hebben voor een andere visie op de belastbaarheid.
De diagnose kan iets zeggen over de prognose.

voorbeeld De behandelaar kan iemand volledig arbeidsongeschikt vinden of blijvend ongeschikt voor rugbelastend werk, van belang is te weten of dit op valide argumenten berust.
De prognose bij een metastase in de rug is anders dan bij artrose.

Onderzoek

Onderzoek lichamelijk en psychisch:

Naast regulier lichamelijk en psychisch onderzoek dient men ook te letten op het ziektegedrag: bijvoorbeeld hoe staat de cliënt op van de stoel, hoe kleedt de cliënt zich aan, zijn de bevindingen consistent, hoe is het looptempo en -patroon, hoe zit de cliënt op de stoel, geeft de cliënt een krachtige hand?
Psychisch: naast regulier psychiatrisch onderzoek: geeft de cliënt adequate antwoorden? Kan hij de lijn van het gesprek vasthouden? Emotionele stabiliteit? Gevoelens van de onderzoeker bij de cliënt?

Informatie van de behandelaars:

Opvragen van informatie is ter aanvulling van zelf verricht onderzoek en kan bijdragen aan het beoordelen van de 'externe consistentie'. Opvragen van informatie zal niet altijd bijdragen aan het vaststellen van de beperkingen omdat de informatie van behandelaars vaak is gericht op de curatieve invalshoek zoals diagnose en nut van behandeling. Een diagnose is voor een verzekeringsgeneeskundige beoordeling niet beslist noodzakelijk. Verder is de cliënt zelf vaak goed op de hoogte van de bevindingen van de behandelaars. Zie verder ook de standaard "Communicatie met behandelaars".[17]
Indicaties voor het opvragen van informatie zijn onder andere:
- Als onduidelijk is wat er medisch aan de hand is.
- Op verzoek van de cliënt.
- Als de cliënt aangeeft dat er verschil in visie is tussen behandelaar en verzekeringsarts.
- Inconsistenties bij de beoordeling.
- Weinig ziekte-inzicht, (dis)simuleren.

Aanvullend onderzoek:

Er zijn diverse mogelijkheden tot ander aanvullend onderzoek zoals Functional Capacity Evaluation (FCE), psychologische testen, hetero-anamnese, het verrichten van een expertise of arbeidsdeskundig onderzoek en het afnemen van vragenlijsten. Zie ook de standaard 'Onderzoeksmethoden'[18].

Ervaren Belemmeringen

Na een inventarisatie van de klachten dient gevraagd te worden naar de belemmeringen die de cliënt ondervindt vooral in relatie tot arbeid. Ook hier kan bij multiple klachten het beste gestructureerd uitgevraagd worden welke belemmeringen de cliënt ervaart. Er dient daarbij een beeld gevormd te worden van de ernst van de belemmeringen. Het vragen naar concrete gedetailleerde voorbeelden van de belemmeringen geeft vaak een goed beeld van die ernst[19] (Bijlage 1). Cliënt geeft in die voorbeelden namelijk vaak de grens aan van wat hij niet meer aankan. Het is van belang de cliënt zo veel mogelijk zelf die voorbeelden te laten bedenken omdat dit een indicatie geeft van de ernst en de impact op diens leven. Als de cliënt zelf nauwelijks dagelijkse voorbeelden kan bedenken, dan is het twijfelachtig of het ernstige beperkingen zijn.

waarom Iedere arts heeft zijn eigen specialisme en daarbij behorende speciële anamnese. Een cardioloog zal vragen naar klachten op het cardiologisch gebied en een neuroloog op neurologisch gebied. De verzekeringsarts dient de beperkingen vast te stellen op het gebied van arbeid. Zijn speciële anamnese bestaat uit vragen naar de beperkingen en activiteiten van de cliënt.

voorbeeld Bij concentratieproblemen vertelt de ene cliënt problemen te hebben met een moeilijke studie of met doorlezen van beleidsstukken, terwijl een andere cliënt aangeeft dat het doorlezen van een eenvoudig tijdschrift of een soap op de tv volgen al een probleem is.
Bij rugklachten vertelt de ene cliënt dat het tillen van een aktetas niet goed lukt en een andere cliënt krijgt klachten bij het sjouwen van grindtegels.

Niet alle belemmeringen worden spontaan genoemd het is daarom van belang ook navraag te doen naar belemmeringen die vaak optreden bij dat ziektebeeld. Desnoods kan het belastbaarheidspatroon erbij gepakt worden en kunnen punt voor punt de beperkingen besproken worden. Dit geldt voor lichamelijke klachten en psychische klachten. Een manier om de psychische belemmeringen uit te vragen en geen belemmeringen te vergeten is, naast het uitvragen van de items van de FML[20], het uitvragen van de items van de Psychische mogelijkheden lijst (PML).[21,22] Zie ook bijlage 2 en 3.

voorbeeld Bij rugklachten worden vaak problemen met bukken en tillen spontaan genoemd. Zitten, staan en lopen zijn ook regelmatig beperkt en dienen dan ook in de anamnese aan bod te komen.
Bij psychische problemen worden vaak de klachten aangegeven (slaapstoornissen, stemmingsproblemen). De beperkingen zijn echter van belang. Vele psychische beperkingen worden spontaan niet genoemd en zijn vaak wel aanwezig b.v. problemen om structuur te houden, niet aankunnen van emotionele belasting, conflicten en concentratieproblemen.

Participatie

Doel is een beeld te krijgen van de activiteiten die de cliënt ondanks de klachten nog wel onderneemt en welke niet meer. Als de cliënt zaken zelf niet meer doet wie helpt en hoe vaak? Daarbij kan gevraagd worden naar de volgende activiteiten:

- ADL, zelfverzorging
- het regelen van eigen administratie en financiën
- verzorging en opvoeding van kinderen
- huishouden, koken, boodschappen, klusjes in en om huis
- hobby's en vrijetijdsbesteding
- sporten
- werk

- sociale contacten en functioneren in gezin, familie en met andere mensen
- dagverhaal

Ook hier geldt dat het van belang is door te vragen naar concrete en gedetailleerde activiteiten (bijlage 1). Bij het onderdeel 'Ervaren belemmeringen' kon de cliënt aangeven wat niet meer lukt door de ziekte, bij dit onderdeel komt naar voren wat de cliënt nog wel onderneemt. Daarnaast wordt duidelijk wat de cliënt niet meer doet.

Dagverhaal:

Onder het dagverhaal wordt verstaan wat cliënt zegt hoe hij zijn dagen doorbrengt. Geprobeerd wordt een zo gedetailleerd mogelijke indruk van een dag van de cliënt te krijgen: Hoe laat gaat de cliënt naar bed en staat hij op? Wat doet de cliënt na het opstaan van uur tot uur? Ligt hij overdag ook op bed of bank? Veel informatie geeft een concrete anamnese van 'de dag van gisteren'[23] van uur tot uur. Dit is echter vrij intensief en kan als een rechercheur overkomen. Vraagt men echter een gemiddelde dag te beschrijven dan is een meer subjectieve invulling van de cliënt mogelijk. Het hangt af van de duidelijkheid van de beperkingen hoe intensief het dagverhaal uitgevraagd dient te worden.

waarom Een dagverhaal legt vaak meer de nadruk op zaken die de cliënt nog wel kan. Dit kan wat tegenwicht bieden aan het alleen maar vragen naar klachten en beperkingen. Verder is het een krachtig hulpmiddel om te zien of het verhaal van de cliënt consistent is, en geeft het vaak een goed beeld van de ernst van de beperkingen. Als arbeidsongeschiktheid op medische gronden wordt overwogen is het ook een belangrijk hulpmiddel om vast te stellen hoe het persoonlijk en sociaal functioneren is.

voorbeeld Een cliënt met MS doet 's morgens licht huishoudelijk werk en moet dan 's middags op bed liggen en kan dan 's avonds weer wat energetisch lichte dingen doen: een beperking in het aantal uren dat de cliënt op een dag kan functioneren lijkt aangewezen. Een cliënt met psychische klachten acht zichzelf slechts belastbaar voor halve dagen maar kan wel hele dagen vullen met huishouding, kinderen verzorgen en daarnaast ook nog bestuurslid van de speeltuinvereniging zijn. Er lijken argumenten aanwezig te zijn om de cliënt toch voor hele dagen belastbaar te achten.

Visie cliënt

Hier kan de cliënt aangeven wat hij zelf vindt van zijn beperkingen en mogelijkheden. Duidelijk dient te worden waarom de cliënt vindt dat hij zijn eigen werk niet kan doen, en wat hij vindt van zijn mogelijkheden in ander (lichter) werk. De verzekeringsarts moet hier wel duidelijk maken dat het gaat om beperkingen voor werk in het algemeen (dus niet alleen specifiek voor eigen werk) en dat bij een arbeidsongeschiktheidbeoordeling geen rekening gehouden wordt met de belasting in de thuissituatie. Hier kan de verzekeringsarts ook alvast zijn voorlopige indruk van de belastbaarheid aangeven.

waarom Door deze terugkoppeling kunnen nieuwe beperkingen of nuancering van beperkingen aan bod komen. De cliënt kan argumenten naar voren brengen waarom hij het eventueel niet eens is met de voorlopige visie van de verzekeringsarts.

voorbeeld Cliënt acht zich voor halve dagen belastbaar omdat hij zijn eigen werk maar halve dagen vol kan houden. Hij heeft psychische klachten en kan zich moeilijk concentreren, zijn werk is corrector. De verzekeringsarts geeft aan dat hij de cliënt hele dagen belastbaar acht mits het werk niet gepaard gaat met intensief concentreren. Na uitleg acht de cliënt zich ook wel hele dagen geschikt als b.v. brugwachter of portier waarbij geen intensief concentreren noodzakelijk is.

Visie verzekeringsarts

De verzekeringsarts beoordeelt op basis van de verkregen gegevens of nog aanvullende informatie noodzakelijk is en beoordeelt de belastbaarheid conform wet- en regelgeving, standaarden en protocollen. Beoordelen is niet opschrijven wat de cliënt zegt dat hij kan, maar waar de verzekeringsarts van overtuigd is wat de cliënt nog kan. Van belang zijn onder andere consistentie en plausibiliteit. Het oordeel wordt de cliënt duidelijk een aansluitend bij zijn vocabulaire medegedeeld. Als nog geen oordeel mogelijk is wordt afgesproken hoe en wanneer dit zal gebeuren.

Voorbeeld Er is geen consistentie tussen stoornissen, beperkingen en handicaps als de cliënt aangeeft dat hij veel pijn in zijn schouder heeft en er niks mee kan tillen maar nog wel aan motorcross doet. Er is binnen het domein beperkingen geen consistentie als de cliënt aangeeft dat een volle emmer water wel getild kan worden maar een pak suiker niet. Er volgen twee voorbeelden van de oordeelsvorming op de items zitten en concentreren.

Zitten

Cliënt vertelt:	Ik kan maximaal ¼ uur zitten i.v.m. rugklachten.
Belemmering:	Pijn in de rug bij zitten in de kerk en bij een verjaardagvisite.
Participatie:	Autorijden van Groningen naar Amsterdam (200 km) met een pauze in Zwolle. Kerk- en bioscoop bezoek vindt wel plaats.
Onderzoek:	Zit ½ uur zonder te verzitten zonder ondersteuning op de stoel.
Diagnose:	Chronische aspecifieke rugklachten.

Conclusie van de verzekeringsarts: Zitten kan 1 uur.

Argumentatie: Cliënt kan 1 uur gefixeerd in de auto zitten, dit is weliswaar niet de hele dag, maar ook langer zitten zoals kerk en bioscoopbezoek wordt niet vermeden. Ook tijdens het onderzoek geen aanwijzingen dat de cliënt korter belastbaar is. Ook gezien de diagnose moet dit mogelijk zijn.

Concentreren

Cliënt vertelt:	Ik kan me absoluut niet concentreren i.v.m. burn out.
Belemmering:	Een Libelle lezen gaat wel aardig, krant lezen alleen de wat meer eenvoudige samenvattingen en koppen, niet de langere stukken. Kan twee bladzijden van een detective achtereen lezen.
Activiteiten:	Leest maximaal ½ uur een tijdschrift. Kijkt een paar uur televisie, kijkt wel naar de "Bold and the Beautiful" maar niet (meer) naar "Columbo".
Onderzoek:	Kan de aandacht goed bij een gesprek van een ½ uur houden. Begrijpt de relatief ingewikkelde uitleg van een arbeidsongeschiktheidsbeoordeling redelijk.
Diagnose:	Surmenage.

Conclusie verzekeringsarts: Kan werk aan waarbij kortdurend specifieke aandacht en concentratie voor een bepaalde taak is vereist.

Argumentatie: Cliënt kan eenvoudige teksten en tv programma's volgen, niet langdurige of ingewikkelde zaken. Dit past bij de diagnose surmenage.

Concrete vragen aan de cliënt:

In het algemeen geldt hier dat met de vraagstelling aangesloten wordt bij het niveau van de cliënt. Sommige vragen zullen door de cliënt moeilijk beantwoord worden i.v.m. psychologische hobbels. Het kan dan nodig zijn meerdere vragen te stellen of hulpvragen in te bouwen. Niet alle vragen zijn direct door de cliënt te begrijpen en vereisen uitleg van te voren (bv. de vraag naar de visie van de cliënt zelf naar zijn belastbaarheid voor gangbaar werk). Gesprektechnische vaardigheden zijn noodzakelijk.

Werk:

- Wat houdt u werk in?
- Hoe lang doet u dit werk al? Is er veel veranderd?
- Wat zijn de zware onderdelen in uw werk? (hoe zou u dit vertellen aan een vriend(in) die het werk wil gaan doen?). Zowel psychisch als lichamelijk.
- Hoe was de sfeer op het werk (met collega's en leiding)?
- Waren er problemen of wrijvingen op het werk?
- Vond/vindt u het werk leuk?

Ziekte

Ziektegeschiedenis:

- Wanneer zijn de klachten begonnen?
- Kunt u eens beschrijven wat er verder is gebeurd tot op heden?

Oorzaak:

- Wat is de diagnose van de huisarts of specialist?
- Bij psychische klachten: Waren er dingen gebeurt of bezig thuis? Wat bent u voor een persoon?

Beloop:

- Hoe lang heeft u al last? Hoe lang had u al last voordat u uitviel?
- Is er de afgelopen maanden nog verbetering opgetreden?

Aard:

- Waar heeft u last van, welke klachten heeft u?
- Bent u verder goed gezond?
- Heeft u in het verleden operaties ondergaan?
- Heeft u hier iets aan overgehouden?

Behandeling:

- Heeft u nog behandeling? Zo ja, welke en hoe vaak?
- Wat is de behandeling tot nu toe geweest?
- Verwacht u nog behandelingen?
- Welke medicatie gebruikt u?
- Heeft u nog afspraken met de specialist of huisarts en wat is de frequentie?
- Wanneer is de laatste afspraak geweest?
- Heeft de behandelaar adviezen gegeven?

Onderzoek:
Hoe staat de cliënt op van de stoel (in de wachtkamer, naar de onderzoekskamer en bij vertrek)?
Zit de cliënt rustig op de stoel, hoelang? Of verzit de cliënt steeds of staat hij tussendoor op?
Wat geeft de cliënt verbaal en non-verbaal aan bij functieonderzoek?
Hoe wordt de tas opgepakt? Hoe wordt een hand gegeven?
Verder lichamelijk onderzoek gericht op het ziektebeeld.

Kan de cliënt de vragen adequaat beantwoorden? Moet hij lang nadenken, is hij veel vergeten? Komt de cliënt alleen? Hoe is de interactie met partner?
Algemeen psychiatrisch onderzoek.

Ervaren belemmeringen:
- Welke dingen kunt u niet meer doen door uw ziekte?
- U hebt last van ….(klacht). Wat betekent dat voor u in het dagelijks leven?
- Wat kunt u er wel en wat kunt u er niet mee?
- Kunt u daar eens een voorbeeld van geven? (Per beperking een concreet voorbeeld, doorvragen).
- Hoe vaak komt het voor? (concreet antwoord vragen: hoe vaak gemiddeld per week, hoe lang duurt het gemiddeld?)
- Als het optreedt wat doet u dan? (concreet antwoord: dus niet: 'ik doe niks', maar 'ik lig op bed te slapen of ik lig op bed te lezen'.)
- Hebt u ook last bij…….? (invullen welke beperkingen volgens de verzekeringsarts met de al genoemde beperkingen kunnen samenhangen bv. zitten bij rugklachten. Zo nodig kan de FML en psychische Mogelijkheden Lijst (bijlage 2) punt voor punt doorgenomen worden.)
- Worden de klachten erger door bepaalde activiteiten?

Participatie:
- Welke hobby's heeft u en welke doet u niet meer?
- Sport u (nog)?
- Wat doet u in het huishouden (deed u voordien meer?)
- Werkt u nog en zo ja wat doet u nog?
- Welke klusjes doet u in huis en welke niet meer?
- Bent u getrouwd/samenwonend en heeft u kinderen (zo ja, hoeveel en hoe oud)?
- Hoe is het contact met gezin, vrienden en kennissen?
- Heeft u hulp voor bepaalde zaken die u niet meer zelf kan doen? Zo ja van wie en hoe vaak?
- Hoe ziet uw dag eruit?
    - Hoe laat staat u op en hoe laat gaat u naar bed?
    - Ligt u overdag op bed of bank?
    - Wat doet u 's ochtend, wat doet u 's middags en wat doet u 's avonds? (concreet)

Visie cliënt:
- Welke onderdelen in uw werk kan u momenteel niet aan?
- Stel u gaat uw werk toch doen wat zou er dan mis gaan?
- Vindt u dat u kan werken in ander werk? Zo ja aan welke voorwaarden moet dit dan voldoen?
- Als ik uw verhaal hoor dan denk ik dat u wel geschikt bent voor…….(voorlopige belastbaarheidsvisie verzekeringsarts)… Vindt u dat ook? Zo neen waarom niet?

Visie verzekeringsarts:

- Als ik alles op een rijtje zet vind ik dat u beperkingen heeft voor ……(meest belangrijke beperkingen in het belastbaarheidspatroon noemen).
- Daarmee zou u wel geschikt zijn voor ……(samenvatten van de mogelijkheden bv. licht rugbelastend werk).
- Dit houdt in dat ……(mogelijke gevolgen voor uitkering)
- De volgende stap is …….. (bijvoorbeeld contact met een arbeidsdeskundige)

Literatuur

1. International classification of functioning, disability and health. Geneva: World Health Organisation, 2001.
2. Boer WEL de, Brenninkmeijer V, Zuidam W. Long-term disability arrangements. A comparative study of assessment and quality control. Hoofddorp: TNO work and employment, 2004.
3. 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;25:501-12.
4. Boot CRL. Sick leave in astma en COPD: the role of the disease, adaptation, work, psychosocial factors and knowledge. Nijmegen: Proefschrift, 2004.
5. Hodiamont PPG, Hodiamont-Kroon AME. Besliskunde in de forensische psychiatrie. Tijdschrift voor psychiatrie 1997;39: 484-9.
6. Boer WEL de, Wijers JHL, Spanjer J, Beijl I van der, Zuidam W, Venema A. Gespreksmodellen in de verzekeringsgeneeskunde. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2006;14:17-23.
7. Schattingsbesluit arbeidsongeschiktheidswetten Den Haag: Ministerie van Sociale Zaken en Werkgelegenheid, 2004.
8. Köhnken G. Statement validity analysis and the detection of the truth. The detection of deception in forensic contexts pp 41-63. New York: Granhag, Pär-Anders, Strömwall, Leif, 2004.
9. 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 and criminological psychology 2005;10:225-45.
10. Vrij A. Criteria-based content analysis: a qualitative review of the first 37 studies. Psychology, Public Policy and Law 2005;11:3-41.
11. Tredouw C, Foster D, Allan A, Cohen A, Wassenaar D. Psychology and Law. Lansdowne, South Africa: Juta and Co Ltd, 2005. Chapter 6 Tredoux C, Pooley S. Credibility Assessment.
12. Lamb ME, Hershkowitz I, Sternberg KI, Esplin PW, Hovav M, Manor T, Yudilevitz L. Effects of investigative utterance types on Israeli children's responses. Int J Behav Dev 1996;19:627-37.
13. Spanjer J. De betrouwbaarheid van WAO beoordelingen. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2001; 9:234-241.
14. Spanjer J, Krol B, Brouwer S, Groothoff JWG. Inter-rater reliability in disability assessment based on a structured interview report. Disability and Rehabilitation 2008;30:1885-90.
15. Is beoordeling op basis van een schriftelijke rapportage betrouwbaar? Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2005:13;105
16. Spanjer J. Professionele prognosestelling verkend. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 1998;2:41-6.
17. Communicatie met behandelaars (standaard). Naslag gevalsbehandeling sociale verzekering. Houten/Diegem: Bohn Stafleu Van Loghum, 1996.
18. Spanjer J. Voorbeelden: een hulpmiddel om de ernst van de beperkingen te beoordelen. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2002;10:109-111.
19. Functionele mogelijkheden lijst. Amsterdam: UWV, 2002.
20. Spanjer J. De praktische concretisering van psychische belastbaarheid. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 1996;5:182-6.
21. De speciële anamnese bij psychische beperkingen. Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 2005:13;277-8.
22. Lomax CL, Brown RG, Howard. Measuring disability in patients with neurodegenerative disease using the 'Yesterday Interview'. Int J Geriatr Psychiatry 2004;19:1058-64.

TOTSTANDKOMING

Na een jaar of zes als verzekeringsarts te hebben gewerkt begon ik artsen die pas in dienst waren in de praktijk te begeleiden. Door de roerige tijden en reorganisaties waren er steeds veel artsen op te leiden (gemiddeld wel zo'n vier per jaar). Deze collega's kregen uitleg over hoe volgens mij het beoordelingsgesprek gevoerd diende te worden. Ik merkte in de discussies met deze verzekeringsartsen in opleiding dat voor mezelf steeds helderder werd wat ik zelf deed in een beoordelingsgesprek en waarom. Na een tijd merkte dat ik steeds hetzelfde aan de artsen moest uitleggen, daarom heb ik het op papier gezet. Zo is de "Belastbaarheid Gerichte Beoordelingsmethode" (BGB) geboren.

Door discussies met vakgenoten werd het mogelijk de diverse onderdelen van de BGB steeds concreter te omschrijven en kwamen er onderdelen bij. Ook discussies met verzekeringsartsen die bijna klaar waren met de opleiding en die ik aan de NSPOH les gaf in de beoordeling van psychische beperkingen, droeg daar aan bij.

Bijstelling vond tevens plaats naar aanleiding van de analyse van veertien video opnames van BGB beoordelingsgesprekken die werden gemaakt in het kader van onderzoek. Toen deze video opnames werden getoond aan 22 verzekeringsartsen uit diverse uitvoeringsinstellingen werd ook gedetailleerd commentaar gegeven. Een deel hiervan is verwerkt in aanpassingen in de BGB.

Tussendoor werd steeds de praktische haalbaarheid getest door de methode toe te passen in de praktijk.

In januari 2001 werd deze methode gekozen als voorkeursmethode bij de beoordelingsgesprekken van de 'niet artsen' van de uitvoeringsinstelling 'Cadans'. Deze 'medewerkers verzekeringsarts' (mva) kregen drie dagen training in de BGB en hun superviserende artsen een dag.

Doordat de artsen de mva-en superviseerden en op basis van hun rapportages de belastbaarheid diende te worden vastgesteld bleek het noodzakelijk vooral het onderdeel 'voorbeelden' verder uit te werken. Bovendien werd ik, doordat ik de BGB moest uitleggen aan niet artsen, gedwongen zaken uit te leggen die een arts vaak automatisch doet. Hierdoor werd ook duidelijk dat een arts automatismen heeft die een objectieve beoordeling kunnen verstoren.

Literatuuronderzoek naar gestandaardiseerde interviews zorgden voor onderbouw en aanscherping van onderdelen, bijvoorbeeld uit de juridische literatuur die ging over de geloofwaardigheid van verklaringen.

Bijlage 1.

Voorbeelden: Een manier om de ernst van de beperkingen te bepalen.

Een cliënt met forse pijn in de rug op basis van een door de orthopeed vastgestelde spondylarthrose kan zijn werk als bouwvakker niet meer doen. Cliënt komt moeilijk lopend binnen. Hij zegt dat hij het advies van de orthopeed heeft gekregen het werk te staken. Hij mag niet meer zwaar tillen. Hij krijgt wekelijks fysiotherapie (paraffinepakkingen en massage). Bij lichamelijk onderzoek van de rug zijn dit de bevindingen:
Houding: lichte scoliose, bekken recht.
Functie: flexie, extensie en rotaties van de rug alle wel mogelijk, de bewegingen gaan wel moeilijk.
Geen aanwijzingen voor radiculaire prikkeling.

De verzekeringsarts wordt gevraagd om aan te geven hoe vaak en hoeveel de cliënt kan tillen.
Is er op dit moment informatie genoeg om de vraag te beantwoorden?
Mijns inziens niet: het is wel aannemelijk dat er forse beperkingen zijn voor tillen, de ernst van de beperking kan echter niet nauwkeurig worden omschreven. Het ziektebeeld of de ernst van de klachten zeggen nog niet veel over de ernst van de beperkingen. Hoe bepaal je de ernst van de beperkingen die iemand ondervindt?

Nadat de cliënt is gevraagd welke beperkingen hij heeft, kan in de anamnese aandacht besteed worden aan de ernst van die beperking. Een goede methode om achter de ernst te komen is het vragen van voorbeelden van de beperking en de activiteiten die de cliënt al of niet nog onderneemt. Cliënten geven in die voorbeelden de grenzen van hun kunnen aan. Cliënt noemt geen voorbeelden van zaken die hij heel makkelijk of heel moeilijk kan, maar zaken die hij nog net wel kan of net niet meer. De verzekeringsarts dient deze grens in zijn beoordeling en belastbaarheidspatroon vast te leggen, dit is dus waardevolle informatie.

Cliënt zegt dat hij slecht kan tillen en dat een kilo of drie maximaal is. Cliënt geeft enkele voorbeelden van het til probleem: Zijn kleinzoon kan hij slechts incidenteel eens optillen. Zijn kleinzoon is 1 jaar en onder incidenteel verstaat hij maximaal 3 maal op een dagdeel. Hij voelt dan meer pijn in zijn rug die de rest van de dag blijft hangen, soms moet hij dan zelfs naar bed. Een ander voorbeeld is dat zijn vrouw het wekelijkse kratje bier moet tillen. Zelfs met twee handen lukt dit niet meer, het schiet dan in zijn rug. Als hij in zijn tuin werkt dan tilt hij niet meer een volle emmer water, maar twee halfvolle emmers: aan elke kant een. Ook dit doet hij niet meer dan een keer of 4 per dag, hij gebruikt vaker de tuinslang dan voorheen.

Bij het vragen naar voorbeelden zijn een aantal zaken van belang. Hieronder zullen enkele aandachtspunten worden besproken, ze zullen worden toegelicht aan de hand van bovenstaand

voorbeeld.

Concreet: Er dient naar concrete voorbeelden gevraagd te worden. Dus niet: ik til maar heel weinig en soms wat meer. Maar: ik til mijn kleinzoon, ik til 2 halfvolle emmers water en ik til geen krat bier. Probeer vage aanduidingen zo veel mogelijk te concretiseren. Voorbeeld: wat is incidenteel optillen? wat is soms? wat is kort?, wat is regelmatig? Er dient concreet uitgevraagd te worden: hoe vaak, hoe veel, hoe lang? Er kunnen anders namelijk interpretatieverschillen bestaan tussen beoordelaar en beoordeelde. In bovenstaand voorbeeld wordt doorgevraagd wat de cliënt verstaat onder incidenteel (3 maal per dagdeel).
Inschattingen van de cliënt zijn niet altijd juist: 5 min staan, 3 kg tillen en 1 km lopen zijn grove inschattingen. Een meer objectieve inschatting kan door de verzekeringsarts gemaakt worden op basis van de door de cliënt genoemde voorbeelden, en hoeft niet noodzakelijkerwijs overeen te komen met de door de cliënt ingeschatte ernst van de beperking. In bovenstaand voorbeeld schat de cliënt de belastbaarheid in op 3 kilo en de verzekeringsarts, op basis van de voorbeelden, op 10 kilo. Dit is het verschil tussen beoordelen en registreren wat de cliënt zegt.
Hoe duidelijker en aannemelijker de beperking is hoe minder er hoeft te worden doorgevraagd. Hoe onduidelijker (de beperking past niet bij de klachten of diagnose, discrepantie tussen subjectieve en objectieve zaken) hoe meer voorbeelden gevraagd worden en hoe concreter wordt doorgevraagd.

Gedetailleerd: Hoe meer details hoe nauwkeuriger de beperking kan worden vastgesteld: hoe oud of hoe zwaar is de kleinzoon?, hoe groot is het krat bier (24 flesjes?), is het dan vol of leeg?, wat is de inhoud van de emmers? Hoe duidelijker en aannemelijker de beperking is hoe minder gedetailleerd er hoeft te worden doorgevraagd.
Niet elk onderdeel van het belastbaarheidspatroon hoeft te worden nagevraagd: soms kan de verzekeringsarts zelf de aan- of afwezigheid en de ernst van de beperking bedenken.
Spontaan: Het meest waardevol is het om de cliënt zelf de voorbeelden te laten bedenken. Je kan er van uit gaan dat als iemand beperkingen heeft, dat hij in het dagelijks leven ook geconfronteerd wordt met deze beperkingen. Cliënt dient dan ook spontaan deze voorbeelden aan te kunnen geven. Als de beperkingen niet ernstig zijn, of zelfs worden gesimuleerd, zal dit moeilijker zijn. Wel is het zo dat hoe langer de cliënt beperkingen heeft, hoe meer hij het heeft geïntegreerd in zijn leven en het niet meer als beperking ziet. Iemand met een sociale fobie, die als beperking heeft dat het onder de mensen verkeren een probleem is, en het leven al jaren hierop heeft aangepast zal niet altijd deze beperking spontaan noemen omdat het min of meer een onderdeel van het dagelijks leven is geworden.
Niet altijd zal een cliënt spontaan een volledige beschrijving geven van alle beperkingen. De beoordelend arts zal navraag moeten doen naar niet genoemde beperkingen. Belangrijke items van het belastbaarheidspatroon kunnen nagevraagd worden. Bij psychische klachten zal de cliënt bijvoorbeeld spontaan noemen dat hij niet tegen verantwoordelijkheid en tijdsdruk kan. Bij navraag

kunnen er echter ook problemen zijn met concentreren en omgang met conflicten.
Beste: Cliënt bedenkt zelf voorbeelden van beperkingen. Geef eens voorbeelden van zaken waar u moeite mee heeft?
Goed: De arts bedenkt de beperking, de cliënt de voorbeelden. En hoe gaat het met concentreren?, noem eens voorbeelden hiervan?
Minder: De arts bedenkt de beperking en de richting aan van de voorbeelden. Hoe gaat het met lezen of tv kijken?
Slecht: De arts bedenkt de voorbeelden. U kunt zeker moeilijk lezen en langer tv kijken?
Recent: Vraag naar recente voorbeelden van beperkingen. We zijn immers geïnteresseerd in hoe nu de beperkingen zijn? Vooral als het al weer wat beter gaat, dan kan de cliënt de neiging hebben om aan te geven hoe erg het eerder was.
Consistentie: Bij de voorbeelden dient er consistentie te zijn tussen de voorbeelden en daarnaast natuurlijk ook met andere zaken zoals bijvoorbeeld het lichamelijk onderzoek en het ziektebeeld. In bovenstaand voorbeeld is er consistentie in het gewicht dat de cliënt nog draagt: een kleinzoon van 1 jaar is ongeveer 10 kilogram, twee halfvolle emmers water ook. Een kratje bier is ongeveer 15 kilogram, dit wordt teveel. Ook de frequentie waarin dit maximum van 10 kg getild wordt is consistent: kleinzoon wordt maximaal 3 maal per dagdeel getild, de emmer een keer of 4 per dag.
Waarom: Het kan nuttig zijn om te weten wat er gebeurt als de cliënt deze beperking overschrijdt. Dit kan nuttig zijn om te onderzoeken of het gedrag is gebaseerd op ziekte: vergelijk bijvoorbeeld "Ik stofzuig niet omdat ik er een hekel aan heb" met "Ik stofzuig niet omdat mijn rug pijn gaat doen". Ook kan het nuttig zijn met betrekking tot therapie advies, vergelijk bijvoorbeeld "Ik til niet meer dan 2 kg omdat ik bang ben pijn in mijn hand te krijgen". met "Ik til niet meer dan 2 kg omdat ik anders pijn in mijn hand krijg". In bovenstaand voorbeeld: Wat gebeurt er als de cliënt toch een kratje bier tilt?

Bijlage 2. Psychische Mogelijkheden Lijst

Omschrijvingen van de beoordelingspunten psychische belastbaarheid

(Op de 8 items dient omcirkeld te worden wat iemand maximaal in staat wordt geacht te kunnen presteren.)

1. STRUCTUUR

1. Aangewezen op zeer gestructureerd werk, vaststaande taken en tijdsindeling, geen onverwachtse opdrachten of zaken. Geen noodzaak of mogelijkheid tot zelf structuur aanbrengen.
2. In staat tot werk met soms wat kleine taken tussendoor, maar verder wel werk met structuur en redelijk vaste indeling van taken en tijd. Geen noodzaak zelf te structureren.
3. In staat tot werk waar goed structuur in aan te brengen is (en wat ook voor een deel wel nodig is anders wordt het chaotisch), en dan nog een aantal onverwachte taken heeft.
4. In staat tot werk met regelmatig onverwachte zaken, voor een deel wel te structureren maar voor het grootste deel niet.
5. In staat tot hectisch werk en veel onverwachte zaken (telefoontjes, mensen die iets vragen) en waar ook niet structuur in aan te brengen is.

2. VERANTWOORDELIJKHEID

1. Aangewezen op werk zonder verantwoordelijkheid, constant iemand die de zaken die mis kunnen gaan controleert, er kan niks mis gaan.
2. In staat tot werk met lage verantwoordelijkheid, er kan niet veel mis gaan of er is regelmatig controle.
3. In staat tot werk met gemiddelde verantwoordelijkheid, geen eindverantwoordelijkheid, fouten hebben geen grote consequenties.
4. In staat tot werk met hoge verantwoordelijkheid, moet belangrijke beslissingen nemen met grote consequenties. Er is echter goede mogelijkheid tot overleg of er is nog iemand boven met eindverantwoordelijkheid.
5. In staat tot werk met zeer hoge verantwoordelijkheid, heeft in belangrijke beslissingen met grote consequenties een eindverantwoordelijkheid.

3. TIJDSDRUK

1. Aangewezen op werk met erg lage tijdsdruk, de taken zijn in rustig tempo makkelijk te doen.
2. In staat tot werk met matige tijdsdruk, voldoende rustmomenten, geen productiepieken.
3. In staat tot werk met gemiddelde tijdsdruk, met een normaal werktempo is er ook wel even tijd voor wat ontspanning.
4. In staat tot werk met hoge tijdsdruk, met behoorlijk tempo doorwerken is het vol te houden, geen of weinig tijd voor ontspanning.
5. In staat tot werk met zeer hoge tijdsdruk, erg krappe tijdschema's, constant hoog tempo van werken noodzakelijk.

4. EMOTIONELE BELASTING

1. Mag niet omgaan met problemen van derden.
2. Kan wel omgaan met problematiek van derden, mits niet ernstig.
3. Kan omgaan met grote ellende (chronische ziekte en dood bv) van derden.

5. CONCENTRATIE

1. Kan de aandacht bij eenvoudig routinematig werk houden, waarbij geen specifieke eisen aan de concentratie wordt gesteld.
2. Kan werk aan waarbij kortdurend specifieke aandacht en concentratie voor een bepaalde taak is vereist.
3. Kan zich langdurig concentreren, mits de materie niet complex is.
4. Kan zich langdurig intensief concentreren op complexe materie.

6. OMGEVING

1. Aangewezen op werkomgeving die rustig is, geen lawaai en overzichtelijk.
2. In staat tot werken in vrij rustige omgeving, nu en dan wat geluiden of iemand die komt of gaat.
3. In staat tot werken in omgeving met regelmatig geluiden of iemand die komt of gaat.
4. In staat tot werken in een vrij onrustige omgeving met veel geluiden.
5. In staat tot werken in een hectische, lawaaierige en onrustige omgeving.

7. CONFLICTEN

1 Kan absoluut niet tegen conflicten.
2. Kan soms wel een gering conflict aan.
3. Kan wel conflicten aan mits niet in direct contact (bv wel per telefoon of op papier), of in direct contact maar dan niet indringende conflicten.
4. Kan in direct contact wel conflicten aan maar niet met agressieve of onredelijke mensen.
5. Kan conflicten aan in direct contact ook met onredelijke of agressieve mensen.

8. OMGANG MET ANDEREN

1. Moet nagenoeg alleen kunnen werken.
2. Kan met kleine vertrouwde groep mensen werken, niet met onbekenden.
3. Kan werken met i.h.a. vaste mensen, daarnaast soms onbekenden.
4. Kan regelmatig omgaan met onbekende mensen, nu en dan wat meer vaste mensen.
5. Kan werken in grote groepen met onbekende mensen of steeds wisselende contacten.

9. OVERIGE BEPERKINGEN

Beperkingen die niet in bovenstaande items zijn onder te brengen.

_____
_____
_____

Beperkingen Arbeidspatroon:
Uren per dag: _____
Dagen per week: _____
Overig: _____

Datum: _____

Arts: _____

Bijlage 3. Speciële anamnese bij psychische beperkingen

Structuur

- Kan cliënt plannen maken (bv in huishouden, boodschappen doen of zaken in het werk)?
- Kan cliënt zich aan die plannen houden?
- Wat gebeurt er als die structuur wordt verbroken/ onverwachtse zaken?
- Zijn er problemen bij het aankunnen van meerdere zaken tegelijk (bv eten koken en kinderen te
- woord staan)?
- Zijn derden noodzakelijk om structuur aan te brengen?

Verantwoordelijkheid

- Neemt cliënt zelfstandig beslissingen en zo ja welke bijvoorbeeld mbt huishouden, kinderen, werk etc.
- Lukt dit ook als de consequenties van een beslissing groot zijn?
- Welke beslissingen kosten moeite?
- Is er controle of overleg nodig bij bepaalde beslissingen?

Tijdsdruk

- Is het een probleem als iets binnen een bepaalde tijd af moet?
- Zijn er tussendoor rustpauzes nodig?
- Als iets sneller af moet (tempo hoog) lukt dit dan wel?
- Kan cliënt meerdere afspraken op een dag aan?

Emotionele belasting

- Kan cliënt problemen van andere aanhoren of zien?
- Kan hij daar ook wel afstand van nemen?

Concentreren

- Lukt het om de aandacht te richten op iets en de aandacht vast te houden op bepaalde zaken?
- Lukt dit ook langer achtereen?
- Lukt dit ook als de stof gecompliceerd is?

Omgeving

- Kan cliënt tegen drukte en onrust in de omgeving?
- Zijn er problemen bij veel of harde geluiden?
- Zijn er problemen in een omgeving met veel mensen?

Conflicten

- Kan cliënt omgaan met conflicten en weerstanden?
- Kan cliënt voor zichzelf opkomen of een mening verdedigen?
- Hoe gaat dit schriftelijk, telefonisch of in direct contact?
- Hoe gaat dit bij 'moeilijke' mensen (agressieve of onredelijke mensen)?

Omgang met anderen

- Kan cliënt omgaan met andere mensen?
- Lukt dit ook als dit onbekenden zijn of in een groep?
- Zijn er problemen met bepaalde mensen?

SHARE dissertations

Graduate School for Health Research SHARE

This thesis is published within the research program **Public Health Research** of the Graduate School for Health Research SHARE. More information regarding the institute and its research can be obtained from our internetsite: www.rug.nl/share

Previous dissertations from the program **Public Health Research**

Koopmans PC (2009) *Recurrence of sickness absence; a longitudinal study*
Supervisor: prof dr JW Groothoff; Co-supervisor: dr CAM Roelen

Chang CMS (2009) *Ageing with joy; the effect of a physical education programme on the well-being of older people*
Supervisors: prof dr JR van Horn, prof dr JW Groothoff, prof dr MA Vreede; Co-supervisor: dr M Stevens

Krokavcová M (2009) *Perceived health status in multiple sclerosis patients*
Supervisor: prof dr JW Groothoff; Co-supervisors: dr JP van Dijk, dr I Nagyová, dr Z Gdovinová, dr LJ Middel

El-Sayed Hussein El-Baz N (2009) *Effect of clinical pathway implementation and patients' characteristics on outcomes of coronary artery bypass graft surgery*
Supervisor: prof dr SA Reijneveld; Co-supervisors: dr LJ Middel, dr JP van Dijk, dr PW Boonstra

Buitenhuis J (2009) *The course of whiplash; its psychological determinants and consequences for work disability*
Supervisors: prof dr JW Groothoff, prof dr PJ de Jong; Co-supervisor: dr JPC Jaspers

Santvoort MM van (2009) *Disability in Europe; policy, social participation and subjective well-being*
Supervisor: prof dr WJA van den Heuvel; Co-supervisors: dr JP van Dijk, dr LJ Middel

Stewart RE (2009) *A multilevel perspective of patients and general practitioners*
Supervisors: prof dr B Meyboom-de Jong, prof dr TAB Snijders, prof dr FM Haaijer-Ruskamp

Jong J de (2009) *The GALM effect study; changes in physical activity, health and fitness of sedentary and underactive older adults aged 55-65*
Supervisor: prof dr EJA Scherder**;** Co-supervisors: dr KAPM Lemmink, dr M Stevens

Buist I (2008) *The GronoRun study; incidence, risk factors, and prevention of injuries in novice and recreational runners*
Supervisors: prof dr RL Diercks, prof dr W van Mechelen; Co-supervisor: dr KAPM Lemmink

Škodová Z (2008) *Coronary heart disease from a psychosocial perspective: socioeconomic and ethnic inequalities among Slovak patients*
Supervisor: prof dr SA Reijneveld; Co-supervisors: dr JP van Dijk, dr I Nagyová, dr LJ Middel, dr M Studencan

Havlíková E (2008) *Fatigue, mood disorders and sleep problems in patients with Parkinson's disease*
Supervisor: prof dr JW Groothoff; Co-supervisors: dr JP van Dijk, dr J Rosenberger, dr Z Gdovinová, dr LJ Middel

Bos EH (2008) *Evaluation of a preventive intervention among hospital workers to reduce physical load*
Supervisor: prof dr JW Groothoff; Co-supervisor: dr B Krol

Wagenmakers R (2008) *Physical activity after total hip arthroplasty*
Supervisors: prof dr S Bulstra, prof dr JW Groothoff; Co-supervisors: dr M Stevens, dr W Zijlstra

Zuurmond RG (2008**)** *The bridging nail in periprosthetic fractures of the hip; incidence, biomechanics, histology and clinical outcomes*
Supervisor: prof dr SK Bulstra; Co-supervisors: dr AD Verburg, dr P Pilot

Wynia K (2008) *The Multiple Sclerosis Impact Profile (MSIP), an ICF-based outcome measure for disability and disability perception in MS: development and psychometric testing*
Supervisors: prof dr SA Reijneveld, prof dr JHA De Keyser; Co-supervisor: dr LJ Middel

Leeuwen RR van (2008) *Towards nursing competencies in spiritual care*
Supervisors: prof dr D Post, prof dr H Jochemsen; Co-supervisor: dr LJ Tiesinga

Vogels AGC (2008) *The identification by Dutch preventive child health care of children with psychosocial problems : do short questionnaires help?*
Supervisors: prof dr SA Reijneveld, prof dr SP Verloove-Vanhorick

Kort NP (2007) *Unicompartmental knee arthroplasty*
Supervisor: prof dr SK Bulstra; Co-supervisors: dr JJAM van Raay, dr AD Verburg

Van den Akker-Scheek I (2007) *Recovery after short-stay total hip and knee arthroplasty; evaluation of a support program and outcome determination*
Supervisors: prof dr JW Groothoff, prof dr SK Bulstra; Co-supervisors: dr M Stevens, dr W Zijlstra

Van der Mei SF (2007) *Social participation after kidney transplantation*
Supervisors: prof dr WJA van den Heuvel, prof dr JW Groothoff, prof dr PE de Jong; Co-supervisor: dr WJ van Son

Khan MM (2007) *Health policy analysis: the case of Pakistan*
Supervisors: prof dr WJA van den Heuvel, prof dr JW Groothoff; Co-supervisor: dr JP van Dijk

Rosenberger J (2006) *Perceived health status after kidney transplantation*
Supervisors: prof dr JW Groothoff, prof dr WJA van den Heuvel; Co-supervisors: dr JP van Dijk, dr R Roland

Šléškova M (2006) *Unemployment and the health of Slovak adolescents*
Supervisors: prof dr SA Reijneveld, prof dr JW Groothoff; Co-supervisors: dr JP van Dijk, dr A Madarasova-Geckova

Dumitrescu L (2006) *Palliative care in Romania*
Supervisor: prof dr WJA van den Heuvel

The B (2006) *Digital radiographic preoperative planning and postoperative monitoring of total hip replacements; techniques, validation and implementation*
Supervisors: prof dr RL Diercks, prof dr JR van Horn; Co-supervisor: dr ir N Verdonschot

Jutte PC (2006) *Spinal tuberculosis, a Dutch perspective; special reference to surgery*
Supervisor: prof dr JR van Horn; Co-supervisors: dr JH van Loenhout-Rooyackers, dr AG Veldhuizen

Leertouwer H (2006) *Het heil van de gezonden zij onze hoogste wet; de geschiedenis van de medische afdeling bij de arbeidsinspectie*
Supervisors: prof dr JW Groothoff, prof dr MJ van Lieburg, prof dr D Post

Jansen DEMC (2006) *Integrated care for intellectual disability and multilpe sclerosis*
Supervisors: prof dr D Post, prof dr JW Groothoff; Co-supervisor: dr B Krol

Van Ham I (2006) *De arbeidssatisfactie van de Nederlandse huisarts*
Supervisors: prof dr J de Haan, prof dr JW Groothoff; Co-supervisor: dr KH Groenier

JansenGJ (2005) *The attitude of nurses towards inpatient aggression in psychiatric care: the development of an instrument*
Supervisors: prof dr SA Reijneveld, prof dr ThWN Dassen; Co-supervisor: dr LJ Middel

Post M (2005) *Return to work in the first year of sickness absence; an evaluation of the Gatekeeper Improvement Act*
Supervisors: prof dr JW Groothoff, prof dr D Post; Co-supervisor: dr B Krol

Landsman-Dijkstra JJA (2005) *Building an effective short health promotion intervention; theory driven development, implementation and evaluation of a body awareness program for chronic a-specific psychosomatic symptoms*
Supervisor: prof dr JW Groothoff; Co-supervisor: dr R van Wijck

Bakker RH (2005) *De samenwerking tussen huisarts en bedrijfsarts*
Supervisor: prof dr JW Groothoff; Co-supervisors: dr B Krol, dr JWJ van der Gulden

Nagyová I (2005) *Self-rated health and quality of life in Slovak rheumatoid arthritis patients*
Supervisor: prof dr WJA van den Heuvel
Co-supervisor: dr JP van Dijk

Dankwoord

Het onderzoek voor dit proefschrift is eigenlijk al begonnen in 2000. Ik was toen als beleidsmedewerker gedetacheerd bij het LISV, waar ik met Wout de Boer prikkelende discussies had over het verzekeringsgeneeskundige beoordelingsgesprek. Wout stond achter "zijn" Methodisch Beoordelingsgesprek, maar stelde zich ook open op voor mijn mening dat het gesprek er anders uit zou moeten zien. Ik werd door hem gestimuleerd om mijn ideeën, die ik tot dusver slechts op een A4-tje had opgeschreven, verder uit te werken en zo is het Belastbaarheid Gerichte Beoordelingsgesprek (BGB) geboren. Ik kreeg zelfs de kans om de bruikbaarheid en betrouwbaarheid van de BGB wetenschappelijk te onderzoeken. De publicatie van dit onderzoek is opgenomen in dit proefschrift. Van onze destijds serieuze plannen om samen promotieonderzoek te doen is niets gekomen, maar het is wel grappig dat we nu elk ons eigen proefschrift mogen verdedigen op ongeveer hetzelfde tijdstip.

De volgende stap werd gezet in 2001 toen een pilot 'medewerker verzekeringsarts' (mva) werd gestart. Niet-artsen ondersteunden de verzekeringsarts door cliënten te interviewen. De pilot was een succes op nagenoeg alle fronten, helaas werden we terug gefloten door de politiek. De BGB werd gekozen als het standaard interview protocol van de mva vanwege de duidelijke structuur. Het intensieve en nadrukkelijke gebruik van de BGB in deze pilot gaf mij een goede mogelijkheid te onderzoeken of deze BGB rapportages betrouwbare beoordelingen opleverden. Een tweede hoofdstuk van het proefschrift was het gevolg. Ik wil daarom allen die hebben meegewerkt aan de mva pilot hartelijk bedanken.

Het moet ongeveer in 2003 geweest zijn dat ik bij Johan Groothoff aanklopte met de vraag of ik bij hem promotieonderzoek kon gaan doen. Hij was enthousiast en wilde me graag als promotor begeleiden, maar omdat "de zak met geld" ontbrak ging dat toen niet door. Later kwam de academisering van de verzekeringsgeneeskunde in snel vaarwater terecht en ontstond de mogelijkheid om promotieonderzoek te verrichten in Amsterdam. Te ver voor iemand die met z'n voeten vastzit in de Groninger klei (wonen in de kop van Drenthe lukt nog net). Gelukkig bracht Jaap van de Gevel, regiomanager UWV Groningen en Drenthe, uitkomst. Jaap draagt de academisering van de verzekeringsgeneeskunde in het Noorden een warm hart toe. Aan hem is te danken dat ik in 2007 echt kon beginnen als parttime promovendus en nog wel als onderzoeker van mijn eigen BGB.

Met Johan Groothoff als promotor en Boudien Krol en Sandra Brouwer als copromotoren ging ik van start. Ik dacht destijds dat ik al een hele onderzoeker was; achteraf bleek echter dat er nog veel te leren was op onderzoeksgebied. Johan wil ik in het bijzonder bedanken voor zijn faciliterende rol, zijn pragmatisme ("het artikel moet deze week de deur uit") en zijn snelle en ter zake doende reacties op concepten. Boudien was een steun in het helpen overwinnen van de koudwatervrees om in het Engels te gaan schrijven en bij het strak formuleren en structureren van artikelen. Sandra heeft tijdens de systematische review veel werk verzet en zette aan tot nadenken, met name tijdens

onze discussies over het wetenschappelijk volledig zijn versus het behouden van de leesbaarheid van een artikel. Boudien en Sandra bedankt.

Omdat ik niet echt was ingebed in een onderzoeksetting zat ik vaak eenzaam in mijn joggingbroek achter de computer op zolder. Gelukkig kon ik met Jan Buitenhuis, die in een vergelijkbare situatie zat, mijn ervaringen delen. Regelmatig bespraken we, onder het genot van een paar witbiertjes met citroen, hoe fout de tijdschriften waren die een artikel van ons afwezen, hoe moeilijk de statistiek en hoe vreselijk belangrijk ons werk toch wel was.

Er zijn veel mensen die een bijdrage aan diverse onderzoeken hebben geleverd. Roel Popping en Sjaak Broersen wil ik bedanken voor de statische ondersteuning. Truus van Ittersum voor de hulp bij de zoektocht naar relevante literatuur. In het bijzonder wil ik alle collega verzekeringsartsen bedanken die hebben meegewerkt. Nardy Tuk, Bart Kuipers, Johan Voogd en Siman Wiersma waren zo dapper hun gesprekken op de video vast te leggen en te laten beoordelen door collega's. Tientallen collega's hebben dagenlang rapportages, video's of cliënten beoordeeld en er waren erbij die zelfs enthousiast meewerkten aan rollenspelletjes. Verder ben ik dank verschuldigd aan managers en teamondersteuners van diverse kantoren.

De leden van de leescommissie, prof. dr. M.H.W. Frings-Dresen, prof. dr. J.H.B. Geertzen en prof. dr. J.H.B.M. Willems wil ik bedanken voor het beoordelen van dit proefschrift.

Klaske, zelf denk je misschien dat je als partner van een promovendus meer steun en betrokkenheid had kunnen bieden. Volgens mij heb ik juist de stimulans gekregen die ik nodig had. Door je neiging de waarde van mijn inspanningen in de vorm van onderzoek en publicaties te relativeren ("zit daar nou iemand op te wachten?") werd ik vaak geprikkeld om er een schepje bovenop te doen. Verder wil ik je bedanken voor je inzet als trainingsactrice bij de BGB-training. Ik herinner me vooral het oefenen van de rollen in de tuin onder de parasol en het nabespreken van de prestaties van mijn collega's op de terugweg in de auto als erg gezellig.

Jammer dat het afgelopen is.

Curriculum vitae

Curriculum vitae

Jerry Spanjer werd geboren op 29 november 1958 in Groningen. Na de studie Geneeskunde ging hij in 1984 werken als militair arts, assistent chirurgie en (na de specialisatie) als huisarts. Vanaf 1989 is hij werkzaam als verzekeringsarts, aanvankelijk bij de Gemeenschappelijke Medische Dienst, later bij de UWV. De specialisatie tot verzekeringsarts werd afgerond in 1995.
Aanvankelijk werkte hij fulltime als uitvoerend verzekeringsarts, maar vanaf 1998 werd uitvoerend werk gecombineerd met andere werkzaamheden op het verzekeringsgeneeskundig vlak. Van 1998 tot 2002 was hij voor 50% gedetacheerd als beleidsmedewerker bij het Landelijk Instituut voor Sociale Verzekeringen, van 2002 tot 2004 als verzekeringsarts bij het Academisch Centrum voor Arbeid en Gezondheid te Groningen, vanaf 2005 als docent en onderzoeker bij de sectie sociale geneeskunde op de afdeling Gezondheidwetenschappen van het Universitair Medisch Centrum Groningen en sinds 2007 is hij part-time bezig met promotieonderzoek.
Jerry is sinds 1997 actief als praktijkopleider en scriptiebegeleider en geeft onderwijs aan verzekeringsartsen (NSPOH en UWV opleidingen) en geneeskunde studenten (Universiteit Groningen).
Jerry is getrouwd met Klaske en samen hebben ze twee kinderen, Tom (18) en Paul (16). Hobby's zijn onder andere mountainbiken, wielrennen en science-fiction (boeken, films en games).