Quality of evaluation of work disability

de Boer, Wout

Quality of evaluation of work diability

Wouter de Boer

Quality of evaluation of work disability

Wouter de Boer

Quality of evaluation of work disability

ACADEMISCH PROEFSCHRIFT

Wout de Boer

The studies in this thesis were carried out at TNO Quality of Life, Hoofddorp, the Netherlands
The studies were financially supported by a grant from the SIG foundation and co-financed by TNO Quality of Life

Cover design and lay out: Jaap van der Plas, TNO
Printing: Drukkerij De Bink, Leiden

ISBN: 978-90-5986-334-7

Quality of evaluation of work disability

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties ingestelde
commissie, in het openbaar te verdedigen in de Agnietenkapel
op donderdag 14 januari 2010, te 10.00 uur
door

Wouter Ernst Lodewijk de Boer

geboren te Kloosterhaar

Promotiecommissie:

Promotores:	Prof. dr. J.H.B.M. Willems
	Prof. dr. P. Donceel
Co-promotor:	Prof. dr. F.J.H. van Dijk
Overige leden:	Prof. dr. J.W. Groothoff
	Dr. C.T.J. Hulshof
	Prof. dr. N.S. Klazinga
	Prof. dr. J.A. Knottnerus
	Dr. H. Kroneman

Faculteit der Geneeskunde

CONTENTS

Ter nagedachtenis aan Henk Herngreen
Uitvinder van de argumentatieve claimbeoordeling.
En voor mij nog veel meer.

Chapter 1

Evaluation of work disability

Work Disability

People engage in society by participating in various groups and activities. For many individuals, work is very important for their health and social participation [1]. Among people that work, some may get sick and stop working temporarily, but people usually recover and start working again. When people return to work they may start by working only part-time or with a reduced work-load. The sick employee with his or her employer most often find solutions for problems that may arise in the resumption of work. In this process, the sick employee follows the sick role: he reports sick, renounces his work duties, and does his utmost to recover [2]. If necessary, those who do not recover and are unable to return to work quickly, receive professional aid from a health service provider, including a physician, physiotherapist, or medical specialist. Alternatively, help may be provided by personnel management, an occupational health physician, or another professional [3]. In the Netherlands and in many other countries, a system of social insurance has been created that regulates exemption from work and protection against wage loss. This system organises and finances support to return to work [4, 5, 6]. This support is offered to, and if necessary, urged upon, the sick employee because of the interest that the group takes in the participation of the individual [7]. According to the legislation, the first objective of the social insurance is to promote participation in work. The explanatory note that accompanies the law on Work and Income according to Labour Capacity (WIA) [8], paragraph 3.1 states:

"Sick and partly disabled employees are expected to do their utmost to return to the labour-market. In turn, these employees can expect to be given the chance to participate in the work process and to get the chance to develop themselves in it. Consequently, the government expects employers to provide these chances. All parties concerned – employees, employers, insurance companies and the UWV – have to exert their strength to promote durable participation of sick and partly disabled employees in the work process. They should give reintegration a maximum chance of succeeding. At the end of the period of wage payment employer and employee together draw up a reintegration report. There they set forth the activities undertaken in order to reach resumption of work. The summary has to convince the benefit agency that employer and employee have done everything possible for reintegration."

A small portion of those that report sick do not recover and are unable to resume work promptly. After one year of absenteeism, approximately one percent of employees are still unable to work. These persons follow the handicapped role [9]: the expectation shifts from complete recovery to partial recovery at most. They are expected to work according to their abilities and to account for their not working. They are released from

obligations they cannot cope with and their pay is continued, but may be adjusted. If all goes well, these individuals and the assisting professionals have done everything that is reasonably possible to help the worker recover and resume work. In this case, the worker's non-participation is not to be blamed on the worker, but rather on his having disabilities. Using this argument, these persons call upon the second objective of the social insurance system: protection from poverty and marginalisation as a consequence of abandoning work participation. People who do not earn enough to support themselves are protected by a benefit, for as much and as long as their incapacity justifies. The explanatory note with the Work and Income according to Labour Capacity act [8], paragraph 3.1 states:

If the reintegration report shows that the employee is fully and permanently disabled for work, he can claim a benefit from the arrangement of this law for income replacement for fully and permanently disabled people. If the reintegration report shows that employer and employee have sufficiently exerted their strength to realize reintegration, but have not succeeded, a partially work-disabled person can claim a benefit on the ground of the arrangement work resumption of this law for people who are partially fit for work.

In the history of Dutch social insurance, these objectives are not new. The legislation has emphasised the objectives of promotion of participation in work and of protection of income in both the Law on Work Accidents (1901) and even more clearly in the Legal Insurance for Work Disability (1967) [7].

Evaluation of work disability

It is not easy to determine who does and does not need the protection of a benefit. The existence of benefits can generate an unintended demand for protection, which results in the second objective pressuring the first objective: protection from marginalization pressures promotion of participation in work. In a system that functions ideally, one can assume that everyone endorses the goal of the system and that all people act to support reintegration optimally. This means that those claiming a benefit would only need to be evaluated minimally. In reality, however, it is difficult to evaluate the sick worker's activities, or his treatment and support. In reality, in the Netherlands and other OECD countries, long term sick leave and work disability are a trap from which people cannot easily escape, with considerable damage to the individual and society [5]. Thus, requests for benefits or support in maintaining or returning to the job are evaluated, the so-called evaluations of work disability. Employees are evaluated on several aspects, including: sickness and handicaps, working capacity and incapacity, prognosis, and recovery-behaviour [10]. Other factors can be part of the evaluation too, such as the contribution of the employer, treatment of the physicians and coaching of the occupational physician.

These principles represent notions that are valid but at the same time they create a number of dilemmas [4, 11, 12]:
- freedom of choice of the individual concerning his recovery and reintegration versus the ground rules that social insurance wields out of solidarity;
- the capacities and incapacities of the individual and the opportunities versus restrictions of the appropriate work for him or her;
- clinging too long to the hope of full recovery and work resumption, and surrendering too early by accepting handicaps;
- diagnosing in a relationship of help versus evaluating disability in a relationship of justifying claims on provisions and benefits;
- the need of the contractor to maintain an efficient system versus the need of the evaluating social insurance physicians and labour experts to thoroughly evaluate the claim and support the claimants.

A benefit could be provided based on the declaration of a claimant and the testimonies of the professionals that have coached him. This is common in the case of short-term absenteeism. In these situations the employee reports sick and the occupational physician advises the employer and employee about possibilities and restrictions in work. For long-lasting absenteeism and permanent disability, the Institute for Employees Benefit Schemes (*Uitvoering Werknemersverzekeringen* or UWV) examines whether everything has been done to facilitate recovery and furthermore whether the potential to earn wages has decreased significantly. SIPs and labour experts evaluate the claim of disability for work. In the Appendix to this thesis this evaluation is described in more detail.

The client, an employee who submits a claim, has an interest in the evaluation being correct: his social and economical position greatly depend on it in the short-term, and often in the long-term as well. The client's health is best served by participating in the labour market, rather than being excluded [1]. Clients have organised themselves into groups, including associations of patients that exchange information and give each other support. Clients are often unsure whether the evaluations are being performed correctly [13].
The quality of the evaluation is also of great importance to the people who execute the evaluation, including the social insurance physicians (SIPs) and labour experts, as it is their responsibility legally. Their societal value and existence depend on the degree to which they perform the evaluations well. Over time, this role has become professionalised, now including a legally recognised specialisation for which post- graduate education, continuous professional development, and guidelines have been created [7, 14]. Many SIPs are unsure of the quality of their work [12], and furthermore, find it difficult to do their work well within the conditions posed by their employer [15, 16]. This is further complicated by the fact that they lack consensus about what quality criteria should be applied during the evaluation process.

They also have differences of opinion with their contractor the Dutch benefit agency UWV [12, 15, 16].
UWV is the organisation that allows or disallows the benefit to the client. UWV strives to implement the laws safeguarding both the financial and the social support for the law. Implementing the law has to provide a certain social justice. Consequently, UWV prescribes requirements for the evaluators and conditions for their work. These have to warrant a legally correct and qualitatively high standard of evaluating work disability. The effectiveness of these requirements is unknown.

All in all, the implementation of the legislation on work disability is a complex matter in which evaluations are a central item. The quality of these evaluations has not yet been established and an actual overview how to evaluate work disability is still lacking. The central subject of this thesis is therefore the quality of the evaluation for work disability. In the literature the terms 'evaluation' and 'assessment' are used interchangeably. Here the former is used, as it suggests a less objective measurement.

Research questions
In this thesis, the following main questions will be addressed:
1. What is the object of the evaluation of work disability?
2. What is to be understood by the quality of the evaluations of work disability?
3. How can the quality of evaluation of work disability be controlled?

To answer these questions, a number of theoretical viewpoints have been used. Six studies have been conducted, relating to the Netherlands and among other countries. As stated, social insurance is not uniquely Dutch. Studying other approaches in different countries can help identify better solutions than those that would be found in the Netherlands alone. In return, Dutch approaches can be valuable in other countries.

Theoretical viewpoints
Evaluation of work disability is not a simple and isolated measurement of some unambiguous characteristics of an individual. They are evaluations of characteristics of people within a legally determined institutional frame after a process of sick leave, treatment, and attempts to get better. The evaluations are a complex whole that can be analysed from different viewpoints. These viewpoints are described in relation to the three research questions of this thesis.

1 Object of the evaluation

Differing views exist about the evaluation of long-term work disability. The legal criterion of work disability is *"being, as a direct and medically determinable result of disease or handicap, unable to earn more than 35% of the earnings a comparable healthy person earns with customary work"* [17]. This definition leaves a lot of room for interpretation. This flexibility is needed to allow tailored evaluations [11], but simultaneously constitutes a threat to a univocal and legally equal application of the law [18]. Many SIPs seem to focus on filling in the Functional Capacity List [12], which leads to the suggestion that functional capacity is the sole subject of the evaluations. SIPs do not agree on this point: some think that the evaluation should encompass much more [12]. Three viewpoints are described in the literature.

One viewpoint is expressed by the Health Council of the Netherlands (HCN), who divided the evaluation of work disability into four tasks [10]. This leads to four objects of evaluation: social-medical history, actual functional capacities, current treatment and coaching, and finally, prognosis. This division encompasses the content of the Guidelines for SIPs. The General Introduction to these Guidelines [10] states that the four tasks together constitute the evaluation of work disability. The relationship among the tasks is not described, however. The evaluation of the social-medical history implies that the client's environment is assessed as far as treatment and coaching are concerned.

A second viewpoint is the handicapped role [9, 19]. The handicapped role describes expectations and obligations between the individual and his environment if support is needed due to long-term sickness and handicap with no prospect of recovery. The individual is expected to rehabilitate and return to work as soon as possible, and his environment is expected to support this process. The handicapped role fits with the opinion that sick leave is a form of behaviour under certain conditions. The handicapped role may provide a way out of the discussion about capacity or incapacity by implying that both are important.

A third viewpoint is found in the International Classification of Functioning and Health (ICF) [20, 21, 22]. The ICF is a classification of related consequences of disease, including: impairments, activity limitations and participation problems. With environmental determinants and personal factors these consequences have a multi-relational connection. The ICF was designed to make international research on consequences of disease comparable. It also drafts a picture of what work disability entails: not just the medical diagnosis but also the relationship between the disabled person, their work and, for example, health care or social insurance. This approach is sometimes termed biopsychosocial [19]. Though ICF is not designed for evaluation of work disability, it can provide a framework within which such an evaluation can occur. In that case, the evaluation includes the disease, therapy, existing impairments, activity limitations, as well as

contribution of the person and the contribution of the environment, such as the company he works for.
All in all, it is not self-evident precisely what is evaluated in the evaluation of work disability. This is problematic when studying the quality of the evaluations.

II Expert definition of quality
Answering the question of quality requires identifying who defines quality. In the case of professional judgments, quality is primarily defined by experts in the field [23, 24, 25, 26]. The Dutch Society of Insurance Medicine (NVVG) claims such a role [27]. Professional discretion is demanded in all situations where relevant grounds for the decision are determined by the situation itself, more than by rules or knowledge. Judges, in their verdicts, call this "the facts and circumstances of this particular case." This discretion can be handled in a qualitatively sufficient fashion by ensuring the competence of the experts who perform the evaluations and by ensuring that they agree [24, 28]. A more direct way of enhancing evaluators' agreement is to do evaluations in committees. This is an application of the mechanism of Spearman and Brown that explains that a group opinion from experts is more likely to be right than the experts' individual opinions [24]. Following this line of thought, the quality of evaluations is found in the degree of agreement between experts, either in individual cases or using guidelines and other practice tools [28]. Thus the criterion is internal and intersubjective rather than external and objective. In the case of social insurance, the experts are not free to define quality as they please; rather, other parties are included in the process. These will be presented and discussed in IIIb.

IIIa Quality model of the evaluation
Evaluations can be analysed, following Donabedian [23], in terms of structure (people and means present for the evaluation), process (actions in the individual case), output (product), and outcome (result) [29]. Such an analysis permits a more precise identification of quality aspects as compared to when evaluations are considered a black box. In the Netherlands this can be filled in as follows:
The structure contains all that is at the disposal of the evaluator. The client puts forward his individual situation and claim. The evaluator puts forward criteria and guidelines for the evaluation, instruments, and his professional knowledge. In the Netherlands, the SIP and labour expert perform the evaluation. There are also written materials, such as the reintegration reports and medical reports.
The process includes the interaction with the client and all other actions needed to complete an evaluation. A large part of that is the disability evaluation interview, if necessary, completed with a medical examination, the testing of functional capacity, consultation with others, and involvement of specialist expertise [30, 31].

Evaluations result in an output, which includes an advice to the contractor of the evaluations, the UWV. This advice is produced in a report form. In the Netherlands, the report of a SIP contains a specification of the client's functional capacities in the so-called Functional Capacity List (FML). A protocol has been designed to check if SIPs comply with the report format. Individual evaluations lead to individual decisions by UWV. Taken together, these decisions lead to an outcome in society. This outcome can be defined in terms of social costs and benefits. Benefits are, for example, labour participation, health, and social legitimacy of claims that are granted or denied. Costs are, for example, benefits paid, transaction costs of UWV, but also a lack of participation and a deterioration of health because of exclusion from work. What is ultimately important and how that is weighted is mostly a matter of public and political debate [32, 33].

IIIb Parties involved in disability evaluation

Disability evaluations are not independent activities between the client and evaluator; rather, they are executed within a network of involved parties. Consequently, the definition of quality cannot be determined solely by the experts [23, 25]. The parties involved can be described in a basic script of evaluation [24]. The basic script model states that a formal evaluation of people by people involves not only a client and evaluator, but also the contractor of the evaluation and an external supervisor. For this thesis the following parties are of interest: the client, the evaluator, the contractor of the evaluation, the legislator, the professional group of evaluators, the supervisor, and the tribunals [29]. All these parties influence the evaluations and all have expectations about the quality of the evaluations.

The client is not simply a passive object, but an active subject who can gain or lose with the evaluation. Furthermore, as a citizen and payer of premiums, the client is part of the society that installed social insurance.

The evaluator has to work according to standards that are not his personal standards, but instead ones that are valid in his profession [26, 28]. The client and evaluator often feel a natural tension during the evaluation between the individual claim and the general rules of the scheme. The evaluator and contractor feel a different tension between professional quality and demands from the administration of the contractor for an efficient process.

The contractor, UWV, is responsible for quality and efficiency of deploying the scheme. The contractor contracts the evaluators and provides means for their work while demanding efficiency.

The tribunals and the medical disciplinary courts check individual cases and make case laws that set standards for later evaluations.

The supervisor, the Ministry of Social Affairs and Employment, checks the system and thus makes touchstones like the Protocol of Social-Medical Actions. The supervisor prevents the client, evaluator and contractor from making deals like the ones made in the 1970s and 1980s when the disability scheme was used to force large numbers of workers to retire [32]. In this

thesis the script model is used to identify the actors that are relevant for quality control [29].

IIIc Sick leave and coaching of sick leave

Disability evaluations occur after a period of sick leave; specifically, in the Netherlands this starts after about 18 months. Part of the evaluation includes an analysis of the way the client's sick leave has been spent, as well as the treatment and coaching that have been applied. Consequently, the evaluations do not stand apart from the social-medical history of the client. Veerman [34] designed a model of the start and continuation of sick leave. Knegt et. al. [3] described the relevant actors of ongoing sick leave. The Guidelines for SIPs implicitly use a model of sick leave for the evaluation of the social-medical history.

Outline of the thesis

To answer the questions, six studies have been conducted, described in the following chapters.

In chapter 2, a study is presented on how the evaluation of work disability is organised and of how quality is controlled in different countries. This is realised using a combination of questionnaires and interviews in fifteen participating countries. The tables show aspects of what is evaluated, how the evaluations are organised, and how quality is ensured. Central to this is the script model, a process model of sick leave and disability, and the Donabedian frame of quality analysis. This study provides material to answer questions 1 and 3.

In chapter 3, a study is presented describing how SIPs reason when evaluating work disability in different countries. This is realised using a case description in focus groups, composed of SIPs, in four countries. The object of the evaluation and the grounds that are legitimate are compared among the four participating countries. The results have been presented to a larger group of SIPs in each country, with questionnaires. Here, the expert definition of quality is central. This study provides material to answer questions 1 and 2.

In chapter 4, a study is presented on how work disability is evaluated in different countries and the use of guidelines therein. With questionnaires, presented to twelve participating countries, the object of evaluation is specified. In those countries that use guidelines, interviews have been conducted concerning the use of these guidelines. The frame of analysis of Donabedian is central in this study, providing material to answer questions 1 and 3.

In chapter 5, a study is presented regarding guidelines that exist in Germany and the Netherlands for the medical part of the evaluation. These guidelines are tested to evaluate the quality of their development. The expert quality definition is central to this study, together with the frame of analysis of Donabedian and the script model. The study provides material to answer questions 2 and 3.

In chapter 6, a study about the disability interview protocols that have been published in the Netherlands is presented. The knowledge and opinions of the people who designed these protocols are based on and compared to the literature. The object of the evaluation, the expert definition of quality, and the Donabedian frame of analysis are central in this study. The study provides material to answer questions 1, 2, and 3.
In chapter 7, a study is presented on the adherence in practice of Dutch SIPs to the interview protocols and their underlying principles. This is a questionnaire study in a selected population of SIPs. Here, too, the object of the evaluation, the expert definition of quality and the Donabedian frame of analysis are central. The study provides material to answer questions 1, 2 and 3.
In chapter 8, the general discussion, conclusions and recommendations are presented.

References

1 Waddell G, Burton K. **Is work good for your health and well-being?** London TSO 2006

2 Parsons T. **The social system**. London Tavistock 1952

3 Knegt R, Besseling JJM, Hoffius R Veerman TJ, Ybema JF. **Scenes in a reintegration process** *[in Dutch]* Amsterdam HSI 2006

4 Stone D. **The disabled state**. London MacMillan 1985

5 Organisation for Economic Cooperation and Development (OECD). **Transforming disability into ability**. Paris OECD 2003

6 Council of Europe. **Assessing disability in Europe**. Strasbourg CoE 2002

7 Boer WEL de, Houwaart E (eds). **Ability evaluated: work disability evaluation in the Netherlands 1901-2005**. *[in Dutch]* Utrecht NVVG 2006

8 **Law on work and income according to working capacity (WIA)** *[in Dutch]* http://lexius.nl/wet-werk-en-inkomen-naar-arbeidsvermogen Visited 4-9-2009

9 Gordon G. **Role theory and illness**. New Haven College and university press 1966

10 Health Council of the Netherlands. **General introduction to the guidelines for social insurance physicians**. *[in Dutch]* 2006/22 Den Haag Gezondheidsraad 2006

11 Teulings CN, van der Veen RJ, Trommel W. **Dilemma's of social security**. *[in Dutch]* Den Haag VUGA 1997

12 Boer WEL de, Steenbeek R. **Report on difficulties in disability evaluation**. *[in Dutch]* Utrecht NVVG 2006

13 Support group ME. **Protocol in practice**. *[in Dutch]* Groningen Steungroep ME 2009

14 Willems JHBM. **Professionalisation of insurance medicine** *[in Dutch]* Den Haag Ministerie van Sociale Zaken en Werkgelegenheid 2000

15 Choy YW. **Report about the consultation of social insurance physicians and labour experts by the department of social- medical affairs.** *[in Dutch, not officially published]* Amsterdam UWV 2005

16 Pelt W van. **Letter to the Minister of Social Affairs and Employment on problems of performance in UWV**. *[in Dutch]* Wageningen 2-6-2009

17 http://lexius.nl/wet-werk-en-inkomen-naar-arbeidsvermogen Article 4 and 5 *[in Dutch]* Visited 4-9-2009

18 Mabbet D, Bolderson H, Hvinden B. **Definitions of disability in Europe: a comparative analysis**. Uxbridge Brunel University 2002

19 Waddell G, Aylward M. **The scientific and conceptual basis of incapacity benefits**. Norwich: TSO 2005

20 World Health Organisation (WHO). **International Classification of Functioning and Health**. WHO Geneva 2001

21 Slebus FG, Sluiter JK, Kuijer PP, Willems JH, Frings-Dresen MH. **Work-ability evaluation: A piece of cake or a hard nut to crack?** *Disab & Rehab*. 2007 **29**: 1295 - 1300

22 Brage S, Donceel P, Falez F. **Development of ICF core set for disability evaluation in social security**. *Disab & Rehab*. 2008 **30**: 1392-6

23 Donabedian A. **The quality of care; how can it be assessed?** *JAMA* 1988 **260**: 1743-8

24 Hofstee WKB. **Principles of evaluation. Methods and ethics of selection, exams and evaluation**. *[in Dutch]* Lisse Swets & Zeitlinger 1999

25 Mul CAM, Verklei HEM, Hazelzet AM, Roos F de. **Quality of formation of judgment**. *[in Dutch]* Hoofddorp TNO 1999
26 Weggeman M. **Managing professionals: just don't!** *[in Dutch]* Scriptum Schiedam 2007
27 NVVG. **Report on quality control in social insurance medicine**. *[in Dutch]* Utrecht NVVG 2003
28 Hofstee WKB. **Promoting intersubjectivity: a recursive-betting model of evaluative judgments**. *Netherlands Journal of Psychology* 2009 **65**: 2-9
29 Boer WEL, Hazelzet AM, Gerven JHAM van. **Manual of evaluation of working capacity**. *[in Dutch]* Hoofddorp TNO 2002
30 Bont AA de, Berendsen L, Boonk MPA, Brink JC van den. **At the doctor's office**. *[in Dutch]* Zoetermeer CTSV 2000
31 Wind H. **Assessment of physical workability: the utility of Functional Capacity Evaluation for insurance physicians**. Amsterdam PhD Thesis Amsterdam University 2007
32 Parlementaire enquêtecommissie uitvoeringsorganen sociale verzekeringen. **Report of parliamentary enquiry on implementing bodies in social insurance**. *[in Dutch]* Den Haag SDU 1993
33 **Healthcare Professionals' Consensus Statement; Statement of Health and Work** www.workingforhealth.gov.uk 4th March 2008
34 Allegro, JT, Veerman TJ. **Sickness Absence**. In: Drenth PJ, Thierry H, Wolff CJ de. (eds) *Work Psychology. Handbook of Work and Organizational Psychology*, 2nd Edn., pp. 121 – 144. United Kingdom Psychology Press Ltd. 1998

Chapter 2:

Organisation of disability evaluation in 15 countries

De Boer WEL[1], Besseling JJM[2], Willems JHBM[3]

[1] Senior researcher, TNO Quality of Life, The Netherlands
[2] Senior researcher TNO Quality of Life, The Netherlands
[3] Professor of insurance medicine, Academic Medical Centre, University of Amsterdam, The Netherlands

Abstract

Background
Legislation, concepts and approaches in social insurance differ between countries. In 2002 and 2003, a comparison of the organisation of work disability evaluation was carried out in 15 countries. The disability evaluation processes in each of these different countries was described. The aim was to determine whether differences in these processes were attributable to differences in legal criteria.

Methods
A script model of evaluations, the handicapped role and the International Classification of Functioning served as tools for analysis. Information on 15 countries was collected by means of desk research, a questionnaire study and interviews.

Results
Evaluation processes show differences in terms of steps involved, use of professional assessors and time consumption. These differences do not correspond to differences in criteria. Legal criteria are formulated in general terms and are fairly similar. Claimants, the law and assessors partly determine the evaluation process. Institutes of Social Insurance, assessors' professional associations, and tribunals also determine the processes. Assessments can be medical, functional or rehabilitational in nature. Medical assessments are universally employed, often in combination with one or both of these other types of assessment. Quality control is incorporated primarily in the process itself.

Conclusion
Criteria for work disability are very similar between countries but their applications differ. Our results can be used for fine-tuning or re-design of current practices. In quality control, there is room for improvement. The types of operationalisation may be helpful in comparing assessments. The script model was found to be a useful tool for elucidating the influence of actors on disability evaluation processes.

INTRODUCTION

Social security is an important focus of political activity in many countries. Today's societies set great store by promoting individual participation in society and reducing dependence on social benefits. In order to achieve these goals, governments have developed and implemented policies containing elements of income support and integration [1], as well as strategies for the prevention of disease and disability. A comparison of practices and effects between one country and another is an important element in the adaptation of policies [2]. As yet, little is known about how the structure of the social security system influences the prevalence of disability pensions [3].

The evaluation of work disability is a major issue within the context of disability policies. Studies of these evaluations tend to focus on sick leave [4]. Much less attention has been devoted to the problem of long-term incapacity for work [5].
The work disability evaluation process is associated with many problems concerning criteria, policy, and implementation [6, 7, 8]. Increasingly, these evaluations have come to be seen as instruments designed to support policies that encourage people with disabilities to work. However, there are a number of problems in this regard [9, 10]. It is difficult to compare the effectiveness of these policies. This is because the criteria for work disability differ from one country to another, as does the way in which they are applied [4, 11]. While the nature and use of work disability assessments has been researched in the Netherlands [12, 13, 14], the assessments themselves are regarded as something of a black box. It is commonly thought that assessors have great latitude in their decisions, as legal definitions are formulated in very general terms [15, 16, 17]. However, assessments are not conducted in a vacuum by the assessor and claimant. As pointed out by Stone [8] and Teulings [18], assessments take place in Institutes of Social Insurance (ISI) which actually organise processes of disability evaluation. These processes, in turn, include their own assessments. This situation has been studied empirically by Mabbett et. al. [17] and by the Council of Europe [19]. It is at the level of ISI that a balance is sought between equity and responsiveness. The uniform application of legal criteria is balanced by assessments of the individual's problems, needs and capacities [19]. Here too, the tenets of the general legal text are translated into terms applicable to a given individual medical case before being put into practice [8].
We compared disability evaluations in 15 different countries, to enhance the body of knowledge in this area. The core research questions used in this comparative study are listed below.

- How are disability evaluations structured in the countries under study?
- Are differences in structure attributable to differences in legal criteria?

This study examines disability evaluations in the context of public programmes designed to protect individuals from a loss of income due to work disability. It focuses specifically on long-term disability.
Our aim was to explore the structure of disability evaluations (see question 1), so we did not simply assume that the law would be passively applied by all those involved. Instead we used a constructivist approach, which made allowance for the translation of legal criteria into administrative and professional actions. Here we focused on the level of Institutes of Social Insurance. Our aim was to describe the process as it is currently structured. We endeavoured to identify the actors and factors that influence the process. One such factor is clearly the legal criteria, but there may well be others.
Following Donabedian [20] we looked at process, structure and output of the evaluation of disability. We adopted a straightforward approach to process description. This involved the use of sequential process steps, commencing at the start of sick leave and concluding at the end of the appeal and reassessment procedure. The aim of this approach was to identify the steps involved in actual assessments, and the actors who implement them. We paid attention to elements of quality control that might be included in the process itself or imposed on top of it [20, 21]. Quality control may aim at administrative quality (time consumption, completeness of dossier), legal quality (accordance with legal requirements) and professional quality (proper use of knowledge and instruments) [22]. We were especially interested in professional quality. The structure of the assessments themselves was studied using a script model of judgment making [22, 23]. Hofstee [22] describes the making of a judgment as more than just a technical process, arguing that it is also a way in which people deal with each other. On this basis, Hofstee determines the roles that key actors have in judgment making in general. This basically concerns the individual who is judged, the professional who makes the judgment, the contractor of the judgments and an authority to guarantee that parties operate in accordance with the rules. This basic approach can be adapted for different situations. According to the extended script model [23], disability evaluations can be expected to involve interactions between a number of parties, apart from claimant and assessor. The first party is the lawmaker (who establishes the legal framework, the criteria and the implementing institution). The second party is the Institute of Social Insurance (ISI hires the professionals needed to assess the claims, supports them and monitors them by means of requirements regarding process and output.) ISIs are responsible for implementing the law on disability for work. These can be independent institutions (as in Italy and the Netherlands), part of the Ministry of Social Affairs (as in the United Kingdom), the municipality (as in Denmark) or the health fund (as in France). The third party is the professional group of assessors (generally health care physicians who apply medical techniques and provide their expertise, as well as setting socio-medical norms.) Fourth are the tribunals (which intervene in the interpretation of legal criteria,

and establish jurisprudence) and fifth an external supervising body (that provides checks and balances in order to ensure that evaluations are conducted properly). This model is depicted in Figure 1. In this study we took the report of the assessors to be the output of the processes.

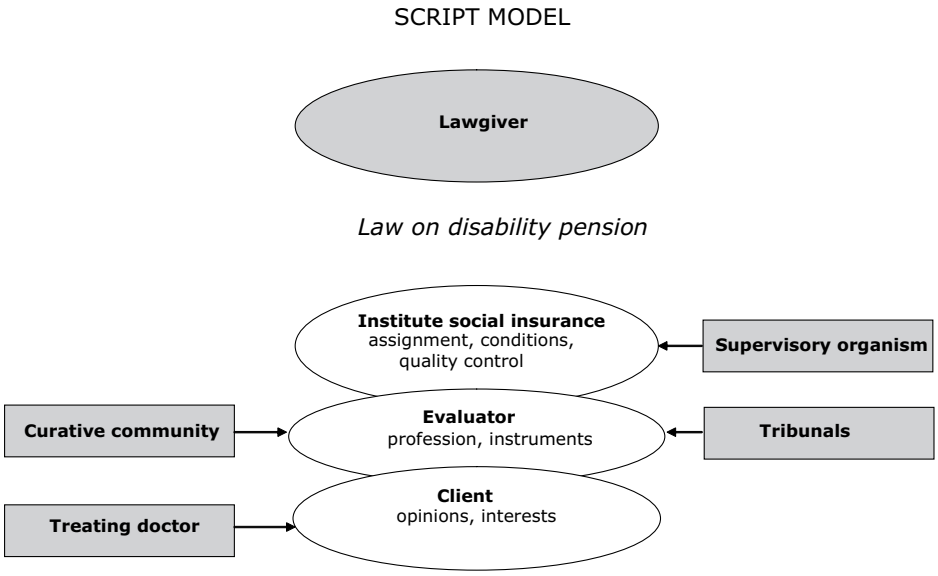

Figure 1. The extended script model

We focused specifically on the relationship between ISIs and assessors. As previously indicated, this relationship is where the legal criteria intended to guide assessors are operationalised [8].

In addressing the second question ("Are differences in structure attributable to differences in legal criteria?"), we focused on those criteria that define the status of disability in terms of the sociological concept of the handicapped role [19, 24]. This concept is an adaptation of the concept of the sick role, taking account of long term disability [25]. The sick role defines the role expectations that exist between the sick individual and people in his social environment. The handicapped defines the role expectations that exist between disabled individuals and those in their social environment. With regard to work disability the handicapped role involves the following issues:

- the claimants' abilities (or inabilities) to do work that one could reasonably expect them to perform;
- health conditions that account for these abilities (or inabilities);
- opportunities and obligations to undergo treatment and rehabilitation.

We omitted non-categorical requirements for obtaining benefits, such as premium history, type of labour contract etc. These are factors that can help individuals to obtain benefits, but they are not involved in the evaluation of the disability itself.
Drawing on experience obtained in previous studies, we explored different phrasings of the criteria in question. An inability to work can be defined as a lack of labour capacity or earning capacity [17]. The health condition may, in accordance with the International Classification of Functioning, Disability and Health (ICF) [26], refer to impairments of body structure and function, anatomical damage (e.g., the loss of body parts), to restrictions of activities (sitting, lifting, concentrating) and to participation (in work and possibly other domains).
In describing our findings, we use the term "assessment" to refer to the assessment by a professional, most often a medical assessor. The term "disability evaluation" refers to the entire processing of the claim up to the final decision, a process that involves other people, including case managers (community employees, usually trained social workers) and administrative staff.

METHODS

The research project was carried out by a group of twelve researchers based in the Netherlands: two from the Dutch Ministry of Social Affairs and Employment, two from the Institute of Social Insurance in the Netherlands (UWV) and eight from the Netherlands Organization for Applied Scientific Research (TNO).

1. Questionnaire and interview

We drafted and fine-tuned a questionnaire by means of an iterative process that involved an examination of the literature, policies and practices with which we were familiar. The questionnaire addressed the following issues:

- The main characteristics of the long-term disability provisions under investigation (name of scheme, position in health field or other precise criteria, etc.)
- The actors involved in the assessment procedure (profession, education, relationships between assessors, etc.)
- The characteristics of the assessment procedure and the process steps involved (input, throughput and output, time schedule, formal power of decision, etc.)
- Quality control (inherent to the process, imposed on the process, measurements, feedback, etc.)

2. Procedure

In each of the states involved, the chief medical advisors of the ISIs head office were responsible for answering or distributing the questionnaire. An ISI generally has one medical doctor who is in charge of the professional work carried out by physicians in the course of assessments. These medical

officers are more intimately involved than anyone else in the issues being addressed in the present study. These individuals were approached through the channels of the European Union of Medicine in Assurance and Social Security (EUMASS) network[1]. They were encouraged to invite others in their organisation to participate, if they felt the individuals in question might be able to contribute to a better understanding of the process. This resulted in the participation of medical and non-medical officers, managers, quality controllers, and other experts.
Questionnaires were sent to and completed in the following countries: Belgium (BE), Denmark (DK), Finland (FI), France (FR), Germany (DE), Great Britain (GB), Hungary (HU), Ireland (IE), Italy (IT), the Netherlands (NL), Norway (NO), the Russian Federation (RU), Slovenia (SI), Spain (ES), and the USA (US).

After filling in and returning their questionnaires, the respondents were interviewed concerning the answers that they had supplied. These semi-structured interviews were conducted in person, by teams of two researchers. The completed questionnaires were submitted to the respondents for final approval.

Interviews of this type were conducted in Belgium, Denmark, Finland, France, Great Britain, Hungary, Italy, the Russian Federation, Slovenia and Spain. The other countries (Germany, Ireland, the Netherlands, Norway and USA) had recently been visited in connection with other research projects related to disability issues. These were studied using the questionnaire alone. The researchers involved in this study contacted all respondents by e-mail in order to check the accuracy of the answers and interpretations that the researchers had noted.

The fieldwork in question took place between September 2002 and April 2003.

3. Data analysis
The data were interpreted in light of the script model, the process model of sick leave and disability, and the handicapped role. This was done by two researchers in an iterative manner: new information was entered while the models were simultaneously checked for completion. In any case of doubt, the original interviewer and/or the respondent were asked to clarify the issue.

[1] The European Union of Medicine in Assurance and Social Security (EUMASS) is a network of associations of medical doctors working in this field.

RESULTS

1. disability evaluation process

a) Process steps

Using information derived from the questionnaires and interviews, we combined all of the steps involved in disability evaluations conducted by the Institutes of Social Insurance (ISI). This made it possible to identify the steps taken in each of the countries involved. The process model is presented in Figure 2. The process entails a period of sick leave, the actual assessment, the formalisation of the decision, and the establishment of an interface with those involved in the rehabilitation process. This involves a total of 11 steps.

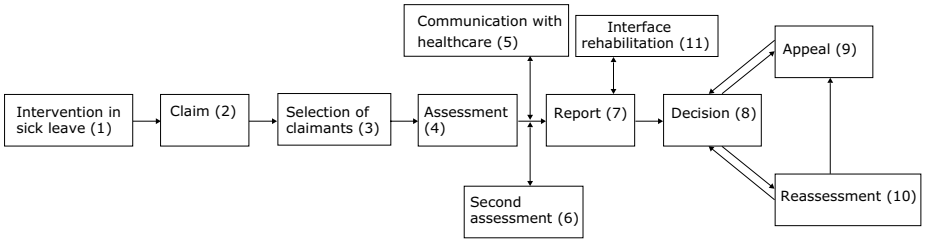

Figure 2. The process steps.

In general, claimants go through a period of sick leave before applying for disability benefits. This 'waiting period' may be of fixed duration or it may be determined in each individual case. We found duration to range from 28 weeks (GB, short term Invalidity Benefit) to five years (DK). Belgium, Ireland, the Netherlands and Great Britain all had a fixed waiting period. The relevant legislation in Hungary and the Russian Federation does not define any waiting period at all.

The period of sick leave can include steps 1 and 2.

Step 1. The ISI may intervene during the period of sick leave. In France, the Medical Advisor must approve treatment and the sick leave after a period of 120 days has elapsed. In Germany and Norway, the Medical Advisor can prescribe rehabilitation before a claim is assessed. In the Netherlands, vocational rehabilitation is examined after 270 days, and must be approved before a claim can be processed. In Denmark, all actions taken by municipal authorities focus on rehabilitation and reintegration; disability will be considered only if rehabilitation efforts fail. ISIs in other countries do not practice any form of intervention.

Step 2. Claiming the benefit. This is generally done by claimants themselves, and only rarely by the attendant physicians (ES, RU). The application form (where applicable) contains details of the claim itself, in combination with information of a medical, social and historical nature. Most of this information concerns treatment and/or rehabilitation. Generally, these claims are submitted to the ISI[2].

The disability assessment may include steps 3, 4, 5 and 6.

Step 3. Claimant selection may be carried out, in order to match claimants with the best qualified assessors (SI, RU) or to award benefits to individuals whose case is so clear-cut that they are exempt from further assessment. This could be done on the basis of disease lists or the severity of the medical condition (GB, IE, NO, US).

Step 4. The actual assessment is carried out by a medical doctor (except in the US). This is often done in combination with other experts. The doctors involved are usually ISI employees. Some work on a case basis for ISI and have their own clinical practice (RU, SI), in GB they work for a private organisation that is contracted by the ISI to perform the assessments.

Method of assessment
We found the following assessment methods in the countries under study:
- assessments conducted in person (FR, HU, IE, IT, NL, RU, SI, ES).
- assessments based on a paper dossier, if necessary completed by an examination conducted in person (BE, FI, DE, GB, NO, US).

In some countries, assessments involve more than one assessor. There may be one or more additional physicians (BE, IT, RU), or a multidisciplinary team (ES, SI). In other countries, assessments are performed by a single ISI physician (FR, IE, NL). If the assessment is inconclusive, extra-information may be gathered. In the United Kingdom the physician works for a private organisation that does assessments for the British ISI.

[2] In Finland, claims are processed simultaneously by a public and a private fund. The two bodies work in close cooperation. This article examines elements from the public processing. Italy's public fund can grant a disability pension (reduction of working capacity over 2/3 and temporary) or an invalidity pension (full loss of working capacity and permanent). If necessary for this article the two will be specified.

In Denmark, evaluation is part of the process of rehabilitation and reintegration. The case manager is the one who controls this process. He consults a medical advisor and other experts when he finds this necessary. In this model, there is no single, distinct assessment. This is because case managers tend to request assessments at various points during the period of rehabilitation and reintegration.

Step 5. Allowance is usually made for some form of communication with health care professionals. In most countries, the attendant physicians are involved in issuing sick leave certificates but play no part in determining the disability pension. In general, health care professionals are only required to provide medical information. They are not required to give their opinion regarding eligibility. In exceptional cases, health care professionals are required to devise the disability evaluation process themselves (NO, RU).

Step 6. Some countries (DE, NL) include an extra-step. This is performed by a non-medical labour expert, who studies the claimant's options on the labour market in the light of the details of their medical report. In the USA, a separate step is taken by the central agency of the ISI. This pre-effectuation review, which is carried out by the central staff of the ISI, involves a check of the accuracy of all decisions to grant a disability pension.

After the assessment, the judgment is formalised. This is a process that can include steps 7, 8, 9 and 10.

Step 7. Drafting a report. In some countries, these reports tend to be brief, containing only the advice if the claimant meets the medical criteria and when he/she is to receive a pension (BE) while in others they can be relatively lengthy, containing a full description of the situation, argumentation and conclusions (NL, DE). In GB and NL, these reports are structured in accordance with standard formats.

Step 8. A formal decision is taken by a social insurance officer (except in BE, FR and RU, where it is the medical advisor who decides the matter, and in DE, where this is done by the labour expert). In practice, the social insurance officer almost always decides in line with the assessor's recommendations.

Step 9. Appeals are always possible, first within the Institute of Social Insurance, and thereafter at a judiciary level.

Step 10. Reassessment is almost always possible (except in IE, IT and NO). Reassessments are basically very similar to primary assessments, although their timing and degree of thoroughness may vary greatly. In Finland, Germany, Italy, the Netherlands (first reassessment), and the USA

a fixed period must elapse before recourse is made to reassessment. In Belgium, France, Hungary, the Netherlands (further reassessments), the Russian Federation, Spain, and the UK, the timing of reassessment may vary.

Finally, in some cases, there may be an eleventh step.

Step 11. In several countries, dealings with rehabilitation structures run parallel to the actual assessments. Claimants may be given a copy of a report containing details of their capacities, together with recommendations regarding their rehabilitation or reintegration. This is done in order to enhance their chances of returning to work. However, as this step falls beyond the scope of our study, we will not discuss it in further detail here.

The duration of the entire process differs greatly from one country to another. The interval between the submission of a claim and the moment when a final decision is made varies from approximately five days (RU) to three or four months (many countries).

The production time of the assessment process, i.e. the amount of time spent on the actual assessment by all the assessors concerned, also exhibits a good deal of variation. Totalling the time required for each of the individual steps (as estimated by the respondents) gives production times that vary from under an hour (RU) to 6 ½ hours (NL).

b) Primary goal of the assessment process
The goal of the assessment may not be limited to checking the claimant's entitlement to disability benefits, it may also include the promotion of their rehabilitation/reintegration. We found verification of disability criteria to be the primary goal of the assessment process in Belgium, Finland, Ireland, Italy, the Netherlands, Norway, Spain, and the USA. The assessment goals listed for Denmark, France, Great Britain, Hungary, Slovenia and the Russian Federation included both verification of entitlement and the promotion of rehabilitation/ reintegration.

c) Actors involved
Having identified the individual steps in the assessment process, we now turn to the actors involved in assessments.

The first actor is, of course, the claimant. He/she is involved in every step of the process, either in person or in the form of their dossier. They claim benefit, support this claim by providing various arguments, and submit to a medical examination. The second actor is the assessor, who is generally a medical advisor, and who often works in conjunction with other assessors. Medical advisors establish a relevant picture of the claimant's state of health and its implications for the claimant's capacity to work.

The Russian Federation, Spain, and Slovenia use committees consisting of several medical advisors. In Belgium, Denmark, Hungary, Italy, Norway and the USA, the number of assessors can depend on the specific case in question. In Denmark and Slovenia, the details of how the assessment is to be organised are decided by a case manager.

In some countries (DE, ES, NL, SI), labour experts are routinely involved. As specialists in labour market conditions and job demands, it is the labour experts' task to establish the relationship between disease, impairment, or functional capacity and participation in labour.

In most countries, it is an administrative officer who formally makes the final decision, not the medical assessor. However, this formal decision is usually in accord with the latter's recommendations.

The ISI determines both the disability evaluation process and the way in which the legal criteria are put into effect (see below). As a result, the ISI is able to influence the process at a very fundamental level.

Assessors are usually medical doctors who are employed by the ISI. Curative health care professionals may be routinely involved in assessments (NO, US, RU) or, depending on the individual case in question, they may be called in at the request of ISI physicians (e.g. IT and NL). Curative health care also has a structural influence, both through professional education and in the establishment of medical norms. In Russia the ISI actively trains attending physicians to be aware of the effect of their treatment and of their professional advice.

Tribunals also have an influence, both at the individual level and at the structural level. The former is exemplified by the way in which their rulings influence the handling of individual cases. We encountered examples of structural influence involving the establishment of case law, the verdicts handed down in appeal cases, and the tribunals charged with the interpretation of the law. This case law is generally incorporated into the instructions that the ISI gives to its assessors. We did not examine the influence of case law in each individual country.

In the script model, the expectation is that an external supervisory body will counterbalance any potential deviations within the ISI. Interestingly, the respondents cited supervisory bodies less frequently than all other sources of influence. In our study, the Netherlands appeared to be the only country to have a supervisory body (the CTSV) with a well defined influence (within a format established for checking reports). Rather than contenting itself with merely issuing this report format, CTSV also insisted that it be routinely used within the Dutch ISI as an internal quality control instrument. The Dutch ISI was required to publish annual reports on

this issue, which CTSV verified by sampling physicians' reports. In other countries, organisations such as the Ministry of Social Affairs or the US General Accounting Office check the statistics either on a regular basis, or by random sampling.

d) Quality control

To some extent, evaluation processes incorporate provisions for quality control. Several methods are included in the process steps described above.

1) Process input (the selection of claimants for appropriate forms of assessment)
2) The process itself (the number and type of assessors can influence the reliability and validity of the assessment)
3) The professionals involved (qualification for the case in question)
4) Process output (report forms can structure the decision and the reasoning)

All of these methods facilitate the task of controlling the quality of the professional (medical and non-medical), legal, and administrative processes.

Provisions for quality control are not limited to the structure of the process itself. External controls may also be imposed. This requires the identification of specific criteria.

The quality criteria, together with the indicators and standards for the quality of the assessment process reported by our respondents, were defined with different levels of precision. Timeliness appears to be clearly defined and monitored. Legal validity is monitored using the results of appeal procedures. However, the quality of decisions (validity and reliability) is poorly defined. In those countries where the monitoring of decision quality does take place, this primarily involves inspections of the individual dossiers (e.g. FR, NL, US). In the Netherlands, the dossiers are inspected using an instrument issued by the external supervisory body. This instrument defines the mandatory items to be included in the report. It also provides some general criteria governing the plausibility and consistency of the findings and reasoning involved.

Other quality control procedures

The respondents cited various other provisions for controlling the quality of assessments. These ranged from consultation between colleagues (peer review) to professional and continuous education, coaching, dedicated forms, protocols, guidelines and other bibliographical references. In some countries, such as Slovenia, appeals are used to improve the quality of assessments. Individual feedback from ISI medical staff to assessors appears to be common practice in the majority of the countries examined. We did not obtain a clear picture of these internal systems for providing

feedback. This may vary from one region to another, and even from one individual to another. In some countries, (BE HU), the assessors' performance is compared to that of other assessors

On the whole, in the course of the interviews our respondents indicated that the quality control of decisions was an issue of growing importance, and one for which definitive solutions have yet to be found.

2. Relationship between the disability assessment process and legal criteria

We explored the relationship between the process of disability evaluation and the legal criteria for disability by examining the way in which these elements are defined. The results are presented in Table 1.

Table 1. Comparison of criteria, operationalisation and disability evaluation

	Belgium	Denmark	Finland	France	Germany	Great Britain	Hungary	Ireland	Italy	Netherlands	Norway	Russian Federation	Slovenia	Spain	United States of America
	BE (IP)	DK	FI	FR	DE	GB	HU	IE	IT	NL	NO	RU	SI	ES	US
Concept of disability	earning capacity	labour capacity	labour capacity	labour capacity + earning capacity	labour capacity	Labour capacity	labour capacity	labour capacity	labour capacity	earning capacity	labour capacity + earning capacity	labour capacity	labour capacity	labour capacity	labour capacity
Waiting time before disability pension	1 year	variable, max 5 years	flexible, max 300 working days	variable, max 3 years	flexible	28 weeks (short term IB); 1 year (long term IB)	undefined	12 months	12 months	52 weeks	flexible	undefined	undefined	flexible, max 18 months	flexible
Levels of disability	1	1	1	2 (+1)	2	1	3	1	1	7(+1)	6	2(+1)	3	4	1
Time schedule reassessment	flexible	no reassessments	no reassessments	flexible, within 3 years	every 3 years	Flexible	flexible or permanent	no reassessments	every 3 years (IP) or permanent (DP)	1st after 1 year, then after 5 years	no reassessments	flexible	Until 45 years every 5 years; over 45 years flexible	2 years	7 years
Operationalisation	medical + functional	medical + functional + rehabilitation	medical	medical + rehabilitation	medical + rehabilitation	Medical + Functional	medical	medical + functional	medical	medical + functional + rehabilitation	medical + rehabilitation	medical	medical + functional	medical + functional	medical + functional

	Belgium	Denmark	Finland	France	Germany	Great Britain	Hungary	Ireland	Italy	Netherlands	Norway	Russian Federation	Slovenia	Spain	United States of America
	BE (IP)	DK	FI	FR	DE	GB	HU	IE	IT	NL	NO	RU	SI	ES	US
Assessors	medical advisor; committee of doctors	case manager; experts as needed: medical, psychological, job consultant	medical advisor	medical advisor	medical advisor, labour expert	adjudicators in social insurance, medical advisor of S. Sema	2 medical advisors	medical advisor	medical advisor, medical specialists	medical advisor, labour expert	general practitioner	3 social insurance clinical specialists	case manager, 2 social insurance clinical specialists, labour expert	medical advisor, labour expert, administrative employee	adjudicator, medical or psychological consultant
Goal of the assessment	check entitlement	check entitlement + promotion rehabilitation	check entitlement	check entitlement + promotion rehabilitation	check entitlement + promotion rehabilitation	check entitlement + promotion rehabilitation	check entitlement + promotion rehabilitation	check entitlement	check entitlement	check entitlement	check entitlement	check entitlement + promotion rehabilitation	check entitlement + promotion rehabilitation	check entitlement	check entitlement
Method of assessment	Paper file, if neces-sary face to face	face to face	paper file	face to face	paper file, if necessary face to face	paper file, if necessary face to face	face to face	face to face	face to face	face to face	face to face	face to face	face to face	face to face	paper file, if neces-sary face to face
Decision taker	medical advisor or committee of doctors	case manager municipality	social insurance officer	medical advisor	labour expert	social insurance officer	social insurance officer	social insurance officer	social insurance officer	social insurance officer	social insurance officer	medical advisors	social insurance officer	provincial director of ISI	social insurance officer

a) Legal definition of disability

As previously stated, definitions of the criteria for disability can vary greatly from one country to another. However, they are broadly based on the following common elements:

- the claimant's ability (or inability) to perform work that one could reasonably expect from a worker in their profession;
- health conditions that account for these abilities (or inabilities);
- opportunities and obligations to undergo treatment / reintegration.

The claimant's ability (or inability) to perform work that one could reasonably expect from a worker in their profession can be described in terms of different concepts. In the regulations that we examined, no reference was made to the claimant's own work, only to work in general (although the exact wording differed from one case to another). According to the ICF, disability can be seen as damage to the organism. This is an approach that focuses on medical signs and symptoms. Disability can also be seen as a characteristic of the individual, by focusing on a restriction of their capacity to function. Alternatively, it can be viewed as a relationship between the individual and society, focusing on a limited capacity to earn money (a criterion of participation). We found differences between the countries in this study in terms of the prevailing concept of disability:

- Earning capacity (Belgium and the Netherlands).
- Combination of labour capacity and earning capacity (France and Norway).
- Labour capacity (the 11 remaining countries).

In none of the programmes in this study damage to the organism itself, or restriction of functional capacity, was used as the sole criterion for determining disability. These criteria are used in combination with the participation criteria: there is always a connection with the claimant's actual work or possibility to work in some type of work. The requirement concerning damage to the organism is introduced by the underlying health condition that accounts for the limitations in question. All legal definitions of disability are couched in terms of damage to health. The exact phrasing varies from one definition to another, but they do not differentiate between health conditions on the grounds of their "acceptability". Examples of phraseology include: "as a direct result of the appearance, or the aggravation, of injuries or functional impairments"(BE); "due to illness, handicap or injury" (FI); "the general state of health, age, physical and mental faculties"(FR); and "any physical or mental impairment" (US).
The least explicit definition of health encountered in this study was the one used by Denmark: "a permanent reduction in the ability to work".
Only in Norway and Spain do the legal criteria of disability make specific mention of opportunities and obligations to undergo professional reintegration. While other countries do have rehabilitation requirements,

they make no mention of these in their legal criteria (DK, FR, DE, NL).
Work disability can be a matter of full or partial disability. It may also be expressed in terms of different levels or categories, indicating partial disability for work (FR, DE, HU, NO, NL, RU, SI, ES). In some countries, an extra level or category is reserved for disabled people who need constant third party assistance (FR, NL, RU).
We attempted to find a relationship between the legal definitions (including the concepts used) and elements of the processes' structure. As can be seen in Table 1, the criteria for a given concept of disability (labour capacity or earning capacity) are not significantly reflected in the structure of the evaluations. Table 1 also illustrates that any combination of structure, criterion, or concept appears to be possible.

b) Operationalisation of disability
The legal definition of disability is formulated in very general terms, so it is open to many different interpretations. There is a need to translate this definition into a more detailed concept, one that can be implemented by the assessors. An examination of the decision, and of its mandatory underlying arguments, reveals three types of operationalisation:
- medical operationalisation, which is characterised by an emphasis on symptoms, diagnoses and impairments. These findings, in themselves, call for decisions regarding disability;
- functional operationalisation, which is characterised by an emphasis on activity (or activity restrictions). These findings lead, either in themselves or through job matching, to decisions regarding disability;
- operationalisation of rehabilitation, which is characterised by an emphasis on the options for rehabilitation, and on previous experience with this approach. These findings also lead to decisions on disability.

Using this typology, we found the following forms of operationalisation in the countries studied here:
- purely medical (FI, HU, IT, RU, US);
- medical combined with functional (IE, SI, ES, GB);
- medical combined with rehabilitation (BE, FR, NO);
- medical combined with functional and rehabilitation (DK, DE, NL).

Medical operationalisation is always involved, often in combination with functional and/or rehabilitative operationalisation.

Table 2 (see next page) presents a comparison of the concept of disability (earning capacity and/or labour capacity) and the way in which the definition is operationalised. As can be seen, almost every country uses labour capacity; this is divided between all forms of operationalisation. Two countries use earning capacity, but operationalise it in different ways. The combined medical and rehabilitative operationalisation includes labour capacity and earning capacity.

Table 2. Relationship between concepts of disability and operationalisation of disability

Operationalisation	Medical	Medical + functional	Medical + rehabilitation	Medical + functional + rehabilitation
Concept				
Labour capacity	Finland, Hungary, Italy, Russian Federation	Ireland, Slovenia, Spain, Great Britain, USA	Germany,	Denmark
Earning capacity			Belgium	Netherlands

DISCUSSION

This article presents the results of a comparative study of disability evaluations in 15 countries. Studies of this type are susceptible to bias. By using a close connection between the models we used and the data obtained, through iteration and by checking our interpretation with the respondents we tried to strengthen the data. We also made detailed comparisons between our results [27] and those obtained by Mabbett et al. [17] and by the Council of Europe [19]. As a result, we feel confident that we have obtained a good picture of the structure of the evaluation process and of the assessments involved, both from the standpoint of the Institute of Social Insurance and that of the assessors.
The picture we obtained is presented from the central perspective of the ISI. As a result, local or regional differences in structure, if existing, are not detected. In addition, our study concentrated primarily on structures, rather than practices. The actual practices used may vary considerably, depending on the nature of the individual professionals [14, 28], claimants [15, 29] and administrative issues involved [13, 15, 29]. The disability evaluations examined by this study took place in late 2002 and early 2003. Though some findings may therefore be outdated by the time this paper is published, we are convinced that the conceptual findings remain valid.

The sequential process model appears to be an effective instrument for describing the processes of disability evaluation with a view to identifying similarities and differences between programmes. The individual processes were found to differ considerably in terms of process steps and duration. In the Russian Federation, three medical specialists are first supplied with extensive details about the claimant from health care sources. These specialists then simultaneously assess the claimant for a period of 15 minutes, which greatly shortens the production time. A report is not asked from them, just a decision. The entire runtime of the process is also very short (five days) as assessors are readily available and no other provisions are needed to produce the evaluations. The final decision emerges from

a discussion between these specialists, during which time only very brief notes are taken. Production times in the Netherlands tend to be very protracted. This is because the physician and the labour expert have to do all the information gathering themselves. They are also required to submit a very extensive report, and to conduct job matching using the national database. The identification of differences in the way that such processes are organised may be of use when countries are thinking about redesigning disability evaluation procedures.
The handicapped role [24] appears to be an effective instrument for comparing legal criteria, as far as disability evaluations are concerned. By this means, we identified many similarities and differences between legal criteria, and in the ways in which they are put into effect. This supports the constructivist approach [15; 16], in which details of organisation and practice are only partly laid down by the law. There are diverse definitions of earning capacity and labour capacity, or of the need for rehabilitation. To say whether this variety corresponds to differences in the actual assessments will require further, more specific study.

The extended script model [23] is useful in identifying influential actors. We were able to obtain a good picture of all those actors whose involvement was predicted, with the exception of external supervisory bodies. We made no attempt to determine whether each of these actors actually possessed the range of influence predicted by the extended script model, as that would have required a longitudinal approach.

ICF [26] is helpful in distinguishing one type of assessment from another. We found three different types (and combinations thereof). This corresponds, in part, to the Council of Europe's typology, particularly with regard to the functional and economic types of assessment. Our three categories offer an impression of the assessment process itself, as they highlight the information gathered and the reasoning presented. The extent to which the typology corresponds to differences in methods and instruments of assessment is still largely a matter of conjecture. It needs to be studied in greater detail.

As noted in the introduction to this article, it is difficult to evaluate the effectiveness of the work disability provisions in each of the countries investigated in this study. Part of the problem is the difficulty inherent in attempts to describe and compare the different countries' disability evaluations. In our view, the models used in this study were helpful in furthering our understanding of these processes. We were able to describe how these evaluations are structured in the various countries. The evaluations were found to differ, they showed certain typologies that led to the legal criteria being operationalised differently in assessment practices. These differences were not fully attributable to differences in legal criteria. If the ways in which these criteria are put into effect (as identified by this

study) are indeed strong determinants of assessment, then this would indicate that social policy is shaped more by the ISI and less by legislators. This conclusion has important implications for governments that, from time to time, seek to steer social policy by redefining legal criteria [2].

Our understanding of the assessments is still hampered by at least two important questions, which we will briefly address here. The first of these is: *What is the relationship between the models used?* In the assessment of disability, an individual connection is made between the legal criteria and the claimant's medical condition. We can define this as a connection between an interpretation of the handicapped role and an interpretation of ICF. These two exhibit areas of overlap, as well as differences. Both make a connection between illness and participation in society but ICF lacks a time perspective, and the grading of normality and disability is unclear in both. It would be interesting to compare assessment procedures, including instruments, in different implementations. This would make it possible to determine whether assessments – at the case level – exhibit differences in medical reasoning and in the use of instruments, and whether these lead to different results. A taxonomy of functional assessment [30] would offer a framework for identifying instruments of assessment. The value of these instruments within social security systems is still open to debate at a fundamental level [31]. This kind of research would foster the development of robust practices of disability evaluation, particularly with respect to quality control [32].
The second question is: *What is the relationship between practice and structure?* Individual assessors deal with individual claimants, each of whom has a specific situation and specific needs. Teulings et al. [18] showed that the application of social security provisions leads to fundamental dilemmas that must be addressed at the individual level. To what extent are rules, process structure and instruments helpful in harmonising equity and responsiveness between claimants? Then there is the element of assessor reliability at the individual and group levels. The rigour of decision-making in this context is a matter of conjecture [14, 33]. This intrinsic problem is compounded by a relational uncertainty. To what extent do the elements of pity and conflict-avoidance dominate individual decision-making [28]?

Given the importance of disability evaluation, and the fact that it is vulnerable to a variety of factors, one might expect certain safeguards to be in place. Possibly in the form of well-established managerial units, either inside or outside the ISIs, that monitor and correct the process of evaluation. By and large, however, quality control appears to be more implicit than explicit and systematic. Our respondents report that the criteria and standards governing the quality of the assessment process are defined in a fairly general terms. The quality of the final decision is

usually monitored only by the inspection of dossiers at the ISI and – on rare occasions – by external parties. Our respondents indicated that several countries were engaged in developing full-blown quality systems. While this is an interesting development, it is by no means sure that it will solve all of the present problems. Even in a relatively refined system, such as that used in the Netherlands, full compliance with, and harmonised application of, the rules and regulations is still difficult [34].

REFERENCES

1 Organisation for Economic Co-operation and Development (OECD). **Transforming Disability into Ability. Policies to Promote Work and Income Security for Disabled People.** Paris OECD 2003

2 Prinz C. **European Disability Pension Policies. 11 Country Trends 1970-2002**. Vienna European Centre 2003

3 Alexanderson K, Norlund A. Swedish Council on Technology Assessment in Health Care (SBU). **Chapter 1. Aim, background, key concepts, regulations, and current statistics**. *Scand J Public Health* Suppl. 2004 **63**: 12-30

4 Allebeck P, Mastekaasa A. Chapter 5. **Risk factors for sick leave - general studies**. *Scand J Public Health* Suppl 2004 **63**: 49-108

5 Aarts L, de Jong PhR. **The Dutch Disability Experience**. In: Prinz C, ed. *European Disability Pension Policies. 11 Country Trends 1970-2002*. Vienna European Centre 2003

6 De Jong PhR. **Disability and Disability Insurance**. In: Prinz C, ed. *European Disability Pension Policies. 11 Country Trends 1970-2002*. Vienna European Centre 2003.

7 Marin B. **Welfare Studies and the OECD Report 2003**. In: Prinz C, ed. *European Disability Pension Policies. 11 Country Trends 1970-2002*. Vienna European Centre 2003

8 Stone D. **The disabled state**. London MacMillan 1985

9 Grammenos S. **Illness, disability and social inclusion**. Dublin European Foundation for the improvement of living and working conditions 2003

10 Waddell G, Aylward M, Sawney Ph. **Back pain, incapacity for work and social security benefits. An international literature review and analysis**. Royal Society of Medicine Press 2003

11 Momm W, Geiecker O. **Disability: Concepts and definitions**. In: *Encyclopedia of Occupational Health and Safety*. Geneva International Labour Organisation 2004

12 Willems JHBM. **Professionalising social insurance medicine: A study on the wishes and possibilities to further professionalise the education for social insurance physicians**. *[in Dutch]* Den Haag Ministry of Social Affairs and Employment 2000

13 Van de Goor AG. **Effects of regulation on disability duration**. Amsterdam Thesis Publishers 1990

14 Kerstholt JH, de Boer WEL, Jansen EJM. **Disability assessments: effects of response mode and experience**. *Disab & rehab* 2006 **28**: 111-115

15 Lipsky M. **Street level bureaucracy**. New York Russel Sage 1980

16 Van der Veen RJ. **Social restrictions of politics**. *[in Dutch]* Leiden Stenfert Kroese 1990

17 Mabbet D, Bolderson H, Hvinden B. **Definitions of disability in Europe: a comparative analysis**. Uxbridge Brunel University 2002

18 Teulings CN, van der Veen RJ, Trommel W. **Dilemma's of social security** *[in Dutch]* Den Haag VUGA 1997

19 Council of Europe. **Assessing disability in Europe**. Strasbourg Council of Europe 2002

20 Donabedian A. **The quality of care; how can it be assessed?** JAMA 1988 **260**: 1743-8

21 Mul CAM, Verkleij HEM, Hazelzet AM, de Roos F. A**n exploratory study into quality control in medical check ups, pre-employment medicals and assessments for disability benefit.** *[in Dutch]* Den Haag Elsevier Bedrijfsinformatie 1999

22 Hofstee WKB. **Principles of assessment. Methods and ethics of selection, examination and evaluation**. *[in Dutch]* Lisse Swets & Zeitlinger 1999
23 De Boer WEL, Hazelzet AM, Gerven JHAM. **Catalogue of evaluations of capacities for work**. *[in Dutch]* Hoofddorp TNO Arbeid 2002
24 Gordon G. **Role theory and illness**. New Haven College and university press 1966
25 Parsons T. **The social system**. Glencoe Free press 1951
26 World Health Organisation (WHO). **International Classification of Functioning, Disability and Health (ICF)**. Geneva WHO 2001
27 De Boer WEL, Brenninkmeijer V, Zuidam W. **Long-term disability arrangements: A comparative study of assessment and quality control**. Hoofddorp TNO 2004
28 Bont A.A., Boonk M.P.A., Berendsen L., Brink J.C.van den. **At the doctor's office: A study of medical assessment of incapacity for work**. *[in Dutch]* Zoetermeer College van Toezicht Sociale Verzekeringen Rapport R00/03 2000
29 Giezen AM, Fiseler JG, Nool JW, Ramakers AF. **Applying the Disability Act by the Social Security Fund for Public Officers**. *[in Dutch]* Zoetermeer CTSV 2001
30 Gaudino EA., Matheson LN, Mael FA. **Development of the Functional Assessment Taxonomy**. *J Occup Rehabil* 2001 **11**: 155-75
31 Matheson LN, Kane M, Rodbard D. **Development of new methods to determine work disability in the United States**. *J Occup Rehabil* 2001a **11**: 143-54
32 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* 2001b **11**: 177-99
33 Razenberg PPA. **Decision making in assessment of incapacity for work. A study into practice**. *[in Dutch]* PhD Thesis. Amsterdam Universiteit van Amsterdam 1992
34 Inspection of Work and Income. **Controlling Management**. *[in Dutch]* Zoetermeer IWI 2004

Chapter 3:
Medico-legal reasoning in disability assessment: a focus group and validation study

WEL de Boer[1], P Donceel [2], S Brage [3], M Rus [4],
JHBM Willems [5]

[1] Department of Quality of Life, TNO, Hoofddorp, the Netherlands
[2] Section of Occupational, Environmental and Insurance Medicine, Department of Public Health, Katholieke Universiteit Leuven, Leuven, Belgium
[3] Department of General Practice and Community Medicine Faculty of Medicine, University of Oslo, Oslo, Norway
[4] Department of assessment, Pension and invalidity insurance of Slovenia, Ljubljana
[5] Coronel Institute, Amsterdam University, Amsterdam, the Netherlands

Abstract

Background
Decisions on disability pensions are based, among others, on medical reports. The way these medical assessments are performed is largely unclear. The aim of the study was to determine which grounds are used by social insurance physicians (SIPs) in these assessments and to determine if the identification of these grounds can help improve the quality of assessments in social insurance practice. The article describes a focus group study and a questionnaire study with SIPs in four different countries.

Method
Using focus group discussions of SIPs discussing the same case in Belgium, the Netherlands, Norway and Slovenia (N=29) we determined the arguments and underlying grounds as used by the SIP's. We used a questionnaire study among other SIPs (N= 60) in the same countries to establish a first validation of these grounds.

Results
Grounds in the focus groups were comparable between the countries studied. The grounds were also recognized by SIPs who had not participated in the focus groups. SIPs agreed most on grounds with regard to the claimant's health condition, and about the claimant's duty to explore rehabilitation and work resumption, but less on accepting permanent incapacity when all options for treatment were exhausted.

Conclusions
Grounds that SIPs use refer to a limited group of key elements of disability evaluation. SIPs interpret disability in social insurance according to the handicapped role and strive at making their evaluation fair trials. ICF is relevant with regard to the health condition and to the process of evaluation. Identification of grounds is a valuable instrument for controlling the quality of disability evaluation. The grounds also appear to be internationally comparable which may enhance scientific study in this area.

Background

In their daily practice, social insurance physicians (SIPs), evaluate claims of incapacity for work. In order to conclude if a claimant is to be accepted as, and compensated for, being disabled, the SIPs always examine file information and often the claimants themselves. This process is governed by legal criteria [1-3] that are formulated in general terms that allow for tailor made decisions. These legal criteria, as far as they pertain to incapacity for work, prove to be quite comparable between countries [1, 3]. All criteria contain, among other things, requirements about a claimant's health condition in relation to work, the permanence of this condition and also about claimant's responsibility to seek therapy and rehabilitation. As such, the criteria are legal representations of the concept of the handicapped role first described by Gordon [4] or the disability role as Waddell & Aylward [5] call it.
The way these legal criteria are implemented varies between countries [3, 6, 7]. The evaluations are carried out in Institutes of Social Insurance (ISI's) that transform the legal assignment into operational categories so as to be able to process claims on a massive scale [8, 9]. This entails operationalisations of the legal criteria, conditions of work (e.g. production time) and prescriptions of work methods (e.g. report forms) [3, 8, 9]. In a previous study by Boer et al. the operationalisations of the legal criteria could be clustered into three categories according to their emphasis on: medical condition (disease, symptoms, impairments), functional status (limitation of activities) and/or required rehabilitative efforts [3]. In practice, operationalisations are frequently combined. In the fifteen countries under study in 2003 Boer et.al. found four different combinations of operationalisations that determine what the SIPs in different countries have to assess: medical only (e.g. Hungary), medical and functional (e.g. Belgium and Slovenia), medical and rehabilitative (e.g. France and Norway), and a combination of all three (e.g. the Netherlands). In many countries assessment of incapacity for work in social insurance is controversial in several aspects: criteria are legal and formulated in general terms offering a large decision latitude [10, 11]; the assessments are carried out in complex organisational settings [12]; the basic concepts of what constitutes disability vary [13]; and the personal encounter between claimant and assessor makes up for conflicts [14, 15, 16, 17].
In this paper we focus on the assessments and their output: the report with arguments and a conclusion. The SIPs who perform the assessments, translate individual claimant situations according to formal ISI requirements and produce reports for the ISI administration. In these reports, information is presented with a conclusion, supported by arguments. The relationship between information, arguments, and conclusions is not univocal. Identical information can lead to different arguments and conclusions in different cases. Toulmin [18] showed that arguments do not simply emerge from information, but rather stand on

the application of more general grounds. With advancing age for example, we can expect physical capacity to decrease (a general ground), and so 'advanced age' can provide an argument in support of incapacity for work that is physically demanding. Advancing age can also be seen as a normal human development, not being a matter of disease (a general ground). In this perspective 'age' can be excluded from the arguments in support of incapacity for work. Consequently, different conclusions may arise in identical cases due to the fact that different grounds are being referred to. If the conclusions of the SIPs are to be legitimate, they need, among others, to refer to grounds that are recognized by all concerned.

Studies in the Netherlands [19, 20] suggest that these grounds can be of different nature: legal (representing the legal criteria), scientific (representing socio-medical evidence), or social (representing social norms as to how to deal with disabled people). Meershoek et. al. [10] found that Dutch SIPs decisions are more social normative than scientific and that this normative dimension need to be transparent if the quality of the assessments is to be guaranteed. This normative aspect is a problem for the SIPs as it is the base of conflicts [21].

Good understanding of argumentation and its grounds can help in developing instruments for assessment, and so help to improve quality of practice. This argumentation oriented approach would be more effective if it were not country specific. We know from earlier research [3] that countries use different operationalisations of the legal criteria. For this reason, we studied the grounds used by SIPs in Belgium, Norway, the Netherlands and Slovenia.

Exploring the grounds in argumentation is relevant to the profession of insurance medicine itself, given the tendency towards internationalization [6], and the priority of claimants' legal security [20, 22-24]. In order to study the arguments and grounds, we can analyse case descriptions. As the grounds are probably implicit, we can use group discussions to make them explicit. Focus group interviews are useful in qualitative research to find common opinions, as they show knowledge and ideas in a context, rather than as individual opinions [25, 26]. In order to establish valid grounds, we need to include SIPs who perform different tasks within the schemes in the countries involved. As the focus group discussions may lead to an agreement on grounds that depend on the group characteristics we need to validate the results with SIPs other than those who participated in the focus group.

Aim

The aim of the study was to determine which grounds are used by SIPs in different countries to support decisions about incapacity for work and to determine if the identification of these grounds can help improve the quality of assessments in social insurance practice.

Method

Subjects

A case of a claim for a disability pension was presented to four focus groups of SIPs working in schemes for long-term incapacity for work in Belgium, the Netherlands, Norway, and Slovenia. The case was selected for presenting a pathology and claimant profile that is met everywhere, and for not being extremely clear and so making any discussion futile. The countries were selected for representing different operationalisations of the legal criteria. The SIPs were selected for their pronounced interest in this project and for their representing different roles or tasks if applicable. To encourage discourse, we aimed at bringing together groups of eight SIPs. We succeeded in bringing together seven SIPs in Belgium, eight in the Netherlands, four in Norway, and eight in Slovenia. Table 1 presents some relevant characteristics of disability pension schemes and of participating SIPs in these countries at the time of our study.

Table 1 Characteristics of disability pension schemes and participants in focus groups in participating countries

Country	Name of scheme	Operationalisation of disability	Time before assessment	Partial incapacity possible	Assess-ment Vis à vis or file or both	SIP's con-cerned (nr included in focus groups)
Belgium	Invalidity Pension	Medical, rehabilita-tional	52 weeks	No	Both	Primary SIP (1), primary ánd secon-dary SIP (6)
Netherlands	WAO (Act on insu-rance of incapacity for work)	Medical, functional, rehabilitational	52 weeks	Yes (7 degrees)	Both	Primary SIP (4), secon-dary SIP (2), appeal SIP (2)
Norway	Disability Benefit scheme	Medical, rehabilita-tional	Not fixed	Yes	File	Primary SIP (3), clinical consultant (1).
Slovenia	Act on Pension and Disability Insurance	Medical, functional	Not fixed	Yes (3 degrees)	Both	Members of primary team (4), members of appeal team (3), external consultant (1)

The case

The participants received a five-page report about a 47-year-old worker in the construction industry, based on a real case of a claim of permanent disability after two years of sick leave. This man had been treated with a lumbar laminectomy and a nerve block. He had osteoarthritis in the right knee. He complained about constant pain and bad sleep. He used medication

for relief of pain and for sleeping. He was obese and moved with difficulty. He was divorced, living alone. He had tried to work as a cab driver but sitting proved to be too heavy. We selected a Dutch case as we knew from earlier research that these had the most elaborate reports. The case was translated into Norwegian and Slovenian. Where necessary, the case was adapted with regard to forms, included organizations, and time perspective to national standards in each country by the researcher from that country. The report consisted of various records: first a record from the occupational health physician on the work and sick leave of the claimant and the efforts to resume work. In this résumé, the medical history, including treatment, was described. Next, the record of the SIP was included, based on an interview with the claimant, a medical examination, and information from the physicians providing medical treatment. The items that were described in the report of the SIP matched the items that SIPs in the Netherlands need to be informed about in order to be able to take a decision on incapacity for work [20]. The items were: the opinion of the claimant on his actual and expected (in) capacities, actual complaints, medical history and general health status, life history, previous work situation, actual private situation, and social situation. The original conclusion was omitted from the case description.

The procedure
We used a stepwise semi-structured approach in which the SIPs initially received the assignment to supplement the case with the information they would normally have in such circumstances, in such a manner that they believed it to be a regular rather than an unusual case. We took note of any proposed alterations, in order to safeguard the medical content of the case. Thereafter, the SIPs were asked to express their opinions before the group session, in terms they would normally use. We noted the conclusions and arguments in this phase as 'primary conclusions'.

Second, the participants were invited to a focus group session. The sessions took about two hours, and were chaired by the leading researcher (WB) and one other researcher. In the sessions, the participants were asked to agree on a typical judgement about the claimant's incapacity for work in line with the legal framework of their own countries. We scored these judgements as 'secondary conclusions'. Participants were free to modify case details as they considered necessary, in order to be able to reach a consensus. We noted these proposed alterations in order to safeguard the medical content of the case.

Third, having reached agreement, they were asked to name arguments that they felt supported or refuted their conclusions. These were listed and agreed upon as being acceptable arguments.
Fourth, during the focus group discussion SIPs were asked to determine the grounds on which these arguments were found to be valid. The grounds

were registered and participants were asked to discuss if these grounds were valid in their normal practice of disability evaluation.

Validation
In a group discussion, the researchers clustered the arguments and grounds produced in the focus groups into four aspects of the assessments. Next, the researchers discussed the grounds in order to identify universal phrasing. Grounds of medical evidence included in focus group arguments were recorded separately.

The grounds as redefined by the researchers were incorporated in a questionnaire that was sent to respondents in each participating country, excluding the participants of the focus groups. We aimed at at least 10 respondents in each country, using pooling that seemed most effective in each country. 20 Belgian SIPs were selected by regional staff SIPs covering the Flemish region. 20 Norwegian SIPs were randomly selected from an existing list of active SIPs in Norway. In the Netherlands SIPs were one by one randomly selected from a list of SIPs that were active in projects of professionalisation. The recruitment went on until 10 respondents had accepted to participate. In Slovenia 10 SIPs were randomly selected from the group of SIPs who worked full time as such; they all responded. Respondents were asked to indicate if the grounds as redefined by the researchers were basically always applicable in the assessment of incapacity for work in their countries.

The agreement with stated grounds by country and total of SIPs was calculated. Differences in agreement were analysed using Pearson's Chi-square test.

Results

The case was recognised as realistic
In all four focus groups, the case was found to be realistic. There were comments from Norway and Belgium about opportunities for rehabilitation and modifications to rehabilitation, and clarifications in the Norwegian group about the region that the claimant lived in. These modifications did not touch on the assessment of the incapacity itself.

Agreement on the degree of disability
Before the focus group sessions, participants were asked to state their opinion of the claimant's disability. In Belgium, seven found him partially incapacitated, and two fully incapacitated. In the Netherlands, seven SIPs found the claimant to be partially incapacitated, and one found him fully incapacitated. In Norway, one found the claimant permanently fully incapacitated, and three found him to be fully but temporarily

incapacitated. In Slovenia, all eight participants found the claimant to be partially incapacitated.
During the sessions, the SIPs were asked to agree on a conclusion for the claimant's disability. After discussion, a common conclusion was reached in all sessions. The Belgian group found the claimant partially incapacitated, capable of performing light unqualified work, and further rehabilitation was suggested. The Dutch group found him partially incapacitated, with restrictions in working in situations of personal risk or of risk to others, working with vibration, doing heavy work, making extreme back movements, and in working in static positions. With these restrictions, the claimant was thought capable of performing full-time work. The Norwegian group found him fully incapacitated for the moment and further rehabilitation was suggested. The Slovenian group found him partially incapacitated, fit for light, quiet work in flexible positions, but limited in walking, crouching, or heavy lifting.

Arguments and grounds
We had asked SIPs to prepare their reasoned arguments before the group sessions. During the sessions, arguments were listed. It was discussed if all arguments were legitimate. Arguments that were proposed by individual SIPs, but rejected by the groups during discussions as not being legitimate were:

Refusing to provide incapacity benefit may make him more ill (No)
The regional labour market is unfavourable (No)
With a pain syndrome he should be active (Si)
Possibly this man wants retirement with a disability pension (Si)
It may be a matter of age (Si)

All other arguments were noted for each country and categorized as being in favour of either (permanent) incapacity or capacity. Thereafter, all arguments were discussed in the focus groups with regard to the grounds they referred to.
In Table 2 (see next page), all grounds are presented together with examples of corresponding arguments. All arguments and grounds are presented in appendix 1.
Some arguments fit with more than one ground. The grounds are clustered around the aspect of disability evaluation that they relate to. We found four aspects, and paid separate attention to the use of medical evidence:

the health condition of the claimant (5 grounds, 16 arguments)
the process of evaluation, a fair trial (6 grounds, 21 arguments)
the time perspective of recovery, treatment and rehabilitation (1 ground, 5 arguments)
the efforts of the claimant to recover and resume work (1 ground, 5 arguments)
medical evidence (4 grounds, 8 arguments)

Table 2: Aspects of disability, grounds, and examples of arguments by country

Grounds	Arguments	Country
Aspect 1. Grounds on claimant's health condition		
It is possible that a health condition is so severe that any form of work is excluded	Clinical and functional impairments are not too severe to prevent him from doing any kind of work	Be, Si
It is possible that a health condition is severe to an extent that it precludes from some work but not all work	His level of functioning is low, too reduced for partial disability.	No
Disability represents a restriction of functional capacities	Several work activities (heavy lifting and carrying etc) are well known risk factors for low back pain and should be avoided	
Capacity for work represents the ability to perform jobs	An unqualified worker can be referred to many jobs, including light work	Be
Advanced age can be a reason to accept restrictions in activities.	His age is not a big problem	Be
Aspect 2. Grounds on a proper process of evaluation		
Findings have to be plausible	There is pathologic evidence of damage of back and knee	Be, Nl, Si, No
Findings have to be consistent	His reduced functioning remains not fully explained with the medical findings	No
Restriction of abilities must not be explained by other factors, notably lack of motivation or opportunity to function	He might comply better with medical advice, is too fat, has a pessimistic view on his future and is inactive	Nl
In order to determine a claimant's abilities his personal experience is a source	His medication causes a lack of alertness. He notices this effect himself.	Nl
In order to determine a claimant's abilities the medical diagnosis is a source	He has pathological degeneration of his back and knees. This risks further damage. Several work activities (heavy lifting and carrying etc) are well known risk factors and should be avoided	Nl, Si, Be,
In order to determine a claimant's abilities the medication is a source	His medication causes a lack of alertness. He notices this effect himself	Nl, Si,
Aspect 3. Grounds on treatment, rehabilitation, and time perspective		
Disability can be accepted as permanent when all treatment options have been tried	He has had all treatment necessary	No
Aspect 4. Grounds on efforts to recover and resume work		
If possibilities for treatment, rehab and/ or work resumption exist the claimant is requested to try these	He might comply better with medical advice, is obese, has a pessimistic view on his future and is inactive.	Nl
Grounds of medical evidence		
Tramadol (an opiate) can cause a lack of alertness	His medication causes a lack of alertness. He notices this effect himself	Be, Nl, Si
Heavy lifting, carrying and the like are well known risk factors for low back pain and should be avoided in the work of people who suffer from low back pain	He has pathological degeneration of his back and knees. This risks further damage. Several work activities (heavy lifting and carrying etc) are well known risk factors and should be avoided	Be, Nl, Si
Chronic low back pain is tiring and may lead to restriction of energetic activities	This kind of chronic pain is tiring and leads to restrictions in energetic activities	Be, Nl, Si, No
Pathologic damage of back and knee make complaints of back and knee plausible	There is pathologic evidence of damage of back and knee	Be, Nl, Si, No

Validation

Questionnaires were returned by 14 SIPs from Belgium, 13 from Norway, and ten from both Slovenia and the Netherlands, resulting in an overall response rate of 78%. Six questions were unanswered (five from Slovenia and one from the Netherlands), out of an expected total of 799 (17 x 47). Results are given in table 3.

Table 3: Agreement with stated grounds, by country and total of social insurance physicians

	Country				Total %
Grounds	Norw N=13	Belg N=14	Neth N=10	Slov N=10	
1 A condition of damaged health can be so severe that any form of work is impossible	12	14	10	10	97,9%
2 A condition of damaged health can be severe to an extent that it precludes from some work but not all work	13	14	10	10	100%
3 Disability is characterised by a restriction of functional capacities	13	12	9	9	91,5%
4 Capacity for work represents the ability to perform jobs	12	12	9	10	91,5%
5 Advanced age is no reason for disability in itself, but can be a reason to accept restrictions in activities	11	14	7	8	85,1%
6 Findings (complaints, symptoms etc) have to be plausible in order to be taken into account in the conclusion	12	10*	10	9	87,2%
7 Findings have to be consistent in order to be taken into account in the conclusion	9	10	10	10	83,0%
8 If restrictions of functional capacities are to be taken into account, they must not solely be explained by other than medical factors, notably lack of motivation or opportunity to function on the labour market	13	12*	10	10	95,7%
9 In order to determine a claimant's functional capacities his personal experience is a source of information	12	12	10	3***	80,4%
10 In order to determine a claimant's functional capacities the medical diagnosis is a source of information	10	14*	9	5**	80,9%
11 In order to determine a claimant's functional capacities the medication is a source of information	8	12	10	7	78,7%
12 Disability can be accepted as permanent when all treatment options have been tried	9	5**	6	10**	63,8%
13 The claimant has the duty to try possibilities for treatment, rehabilitation and/or work resumption when these exist	12	11	9	7	83,0%
14 Tramadol (an opiate) can cause a lack of alertness	10	14*	8	7	83,0%
15 Heavy lifting, carrying and the like are well known risk factors for low back pain and should be avoided in the work of people who suffer from low back pain	8	12	4*	9	70,2%
16 Chronic low back pain is tiring and may lead to restriction of energetic activities	10	10	4*	9	70,2%
17 Pathologic damage of back and knee make complaints of back and knee plausible	8	12	8	8	76,6%

	Country				Total %
Grounds	Norw N=13	Belg N=14	Neth N=10	Slov N=10	
Total grounds 1-5 (Health condition of claimant)	61	66	45	47	93,2%
Total grounds 6-11 (Process of evaluation)	64	70	59	44	84,0%
Ground 12 (Time perspective)	9	5	6	10	63,8%
Ground 13 (Obligation of claimant)	12	11	9	7	83,0%
Ground 14-17 (Medical evidence)	36	48	24	33	75,0%
Agreement in total with all grounds	82,4%	84,0%	84,1%	82,9%	83,4%

Differences in agreement are tested with Pearson Chi-square test. The contrast is subgroup vs. all other cases. *: p<0,05, **: p<0,01, ***: p<0,001 for percentages significantly higher or lower than in the entire group.

The total agreement over all items and all countries was 83.4%. The variation in total agreement between countries was very small, from 82.4% to 84.1%.
The highest agreement was reached with grounds on the health condition of the claimant, grounds 1-5 (93%). For grounds related to the proper process of evaluation, agreement was lower (84%). The least agreement was found for the single ground on permanence of disability (65%*). For compliance and for grounds related to medical evidence, agreement reached 83% and 75%* respectively.
When examining the agreement with individual grounds, larger differences were found, ranging from 100% (ground 2 in all countries) to 30% (ground 9 in Slovenia).
Grounds that were the least agreed upon (below 80% agreement) were grounds 11, 12, 15, 16 and 17. The most controversial grounds (between 33% and 50%) within countries are 12 in Belgium; 12, 15 and 16 in the Netherlands; 11, 15 and 17 in Norway and 9 and 10 in Slovenia.

Discussion

Main findings
In a series of sessions with focus groups in different countries, we studied the reasoning of SIPs in disability evaluation in public schemes for long-term incapacity for work by making arguments and grounds explicit in the case of a construction worker with low back pain. A typical case could be constructed and SIPs in three different countries were able to reach the same agreement on the conclusion of incapacity for work. In Norway, agreement was also reached, but with a different conclusion.

The arguments and grounds that were used in the focus groups were quite comparable between the countries studied. As expected the grounds represent the legal criteria of underlying health condition, compliance with

therapy and rehabilitation, and permanence. In addition, six grounds relate to ubiquitous requirements of a fair trail. In the validation study, all grounds were recognized by SIPs who had not participated in the focus groups. SIPs showed high agreement on grounds with regard to the claimant's health condition, and grounds about the claimant's duty to explore opportunities for treatment, rehabilitation, and/or work resumption. Agreement was less on the ground of accepting permanent incapacity when all options for treatment had been exhausted.

Strengths and weaknesses

A stepwise approach was used to produce these results, asking SIPs in different countries to comment on a specific case and reflect on the grounds they used. The countries were selected on the basis of different operationalisations of the legal criteria. This is not a stable fact as policies change [6, 23]. The way these SIPs were recruited may have led to a selection of SIPs with a higher professional interest than average. This was a deliberate choice in order to produce the most explicit and different grounds. The Norwegian group was smaller than planned, possibly leading to higher degree of uncertainty for the results from Norway. The case was a Dutch case about low back pain. It is possible that at least some grounds are specific to the case, especially the grounds of medical evidence. Further research is needed in order to identify other grounds in other cases.

The grounds we found were clustered by the researchers into five aspects of disability evaluation, congruent with the "handicapped role" and a fair trial. The advantage of "handicapped role" is that it also useful to describe legal criteria [1, 3] and so it seems fit for categorizing grounds. This is not to say that grounds could not have been clustered along other lines, for instance legal grounds next to medical grounds. The clustering into aspects showed that aspects are not mutually exclusive. For instance, the grounds that pertain to the condition of the claimant are connected to the way this condition is examined, and thus with the aspect of a proper process of evaluation. Further research may lead to more refined categories.

In the validation study, other SIPs were asked to recognize the grounds identified in the focus groups. There was a high concordance for most aspects. Quantitative research is needed in order to establish how grounds are used in daily practice.

Other studies

In the literature we found two other studies into grounds that SIPs use in their daily practice [19, 27]. In these Dutch studies, the same procedures were followed and comparable results were found. A validation with a questionnaire was not conducted in either study. We did not follow the categorization into legal, scientific, and social grounds as used in these studies, as we think that the aspects we used here are more precise.

An important model used in disability is the International Classification of Functioning and Health [28]. Slebus et.al [29] found that ICF was only partly considered by Dutch SIPs. In our study, the ICF model could be applicable to the grounds on the health condition that has to be evaluated, but not to the grounds of fair trial, rehabilitation and compliance. A complete evaluation of work disability seems therefore to contain more elements than the ICF provides.

Impact
Studying the reasoning of individual SIPs is possible with the method we used. Differences in their personal convictions can be identified with the analysis of grounds. If our results stand in future research, a potentially important key to understanding disability evaluation has been found: the highly individual evaluations in different national contexts appear to obey rules of a specific image of disability and of legal principles of a fair trial. Being explicit about methods of assessment and about grounds used in particular cases will make the evaluations more transparent and more receptive to quality control. For example, the use of scientific evidence becomes more transparent, which seems to be relevant if we look at the differences of opinion about the effect of Tramadol. Another example is the norm of when to decide that a disability is permanent: in this study it appears to be a norm that is open to individual interpretation.

We studied specific countries because in earlier research they were found to be different in the operationalisations of the legal criteria at the level of the Institution of Social Insurance. Our results do not support that proposition on the level of the assessments. In the assessments in our study, the participants reasoned in line with the concept of the handicapped role.

Conclusions

The identification of grounds that SIPs use helps us to understand the practice of disability evaluation. The grounds guide the translation of information into arguments about legal incapacity for work. It is possible to make these grounds explicit, as they refer to a limited group of key elements of disability evaluation.
Further research is needed to validate and extend our findings, e.g. study on the use of grounds and differences between SIPs, and subgroups of SIPs e.g. full time working vs part time and experienced vs little experienced.
SIPs interpret disability according to a concept that meets legal criteria: the handicapped role. Added to that is the concept of a fair trial. The model of disability, as described in ICF, is applicable to describe the health condition aspect of the "handicapped role". ICF also matters with regard to the process of evaluation where consistency is concerned: the consistency is to be found within and between the categories of the ICF model.
Making the grounds in disability evaluation explicit will enhance the quality of medical reports, as the grounds used can be evaluated in individual cases. In professional practice consensus about these grounds is something to strive for. This would also contribute to the transparency and legitimacy of the disability pension schemes as these schemes are open to constant criticism about their capacity to select the right people for a pension.

Ethics committee: this study was not submitted for ethical approval. The study includes physicians who are not asked to perform specific professional actions for this study but to state and to discuss their professional opinions in general.

Competing interests: The authors declare they have no competing interests. The study was supported with a grant from the SIG Foundation who had no involvement in the study itself or in the decision to submit the paper for publication

Authors' contribution: WEL de Boer designed the study, did the field work and prepared the manuscript. P Donceel, S Brage and M Rus organised and chaired the focus group sessions and led the questionnaire rounds in their respective countries. They also participated in drafting and revising the manuscript. JHBM Willems supervised the project and participated in drafting and revising the manuscript.

References

1 Council of Europe. **Assessing disability in Europe Strasbourg** 2002

2 Mabbet D, Bolderson H, Hvinden B. **Definitions of disability in Europe: a comparative analysis**. Uxbridge Brunel University 2002

3 Boer WEL de, Besseling JJM, Willems JHBM. **Organisation of disability evaluation in 15 countries**. *Pratiques et organisation des soins.* 2007 **38**: 205-217

4 Gordon G. **Role theory and illness**. New Haven College and university press 1966

5 Waddell G, Aylward M. **The scientific and conceptual basis of incapacity benefits**. Norwich TSO 2005

6 Organisation for Economic Co-operation and Development (OECD). **Transforming Disability into Ability. Policies to Promote Work and Income Security for Disabled People**. OECD Paris 2003

7 Rezzvy G Schönfelder W, Øiesvold T, Olstad R, Midré G **Between health care and social security--psychiatric patients and the disability pension system in Norway and Russia**. *BMC Health Serv Res*. 2007 **7**: 128

8 Stone D. **The disabled state**. London MacMillan 1985

9 Lipsky M. **Street level bureaucracy**. New York Russel Sage 1980

10 Meershoek A Krumeich A Vos R. **Judging without criteria? Sickness certification in Dutch disability schemes**. *Sociology of Health and Illness* 2007 **29**: 497-514

11 Manaouil C, Graser M, Jarde O. **Invalidity in the general health scheme**. *Presse Medicale* 2003 **32:** 1622-6

12 Ydreborg B Ekberg K, Nilsson K **Swedish social insurance officers' experiences of difficulties in assessing applications for disability pensions--an interview study** *BMC Public Health*.2007 **7**: 128.

13 Grönvik L **The fuzzy buzz word: conceptualisations of disability in disability research classics**. *Sociology of Health and Illness*. 2007 **29**: 750-66

14 Nüchtern E **The physician as expert evaluator. Aspects of an ambivalent role exemplified by the evaluation of claims in legal health insurance** [in German] *Offentliches Gesundheitswesen* 1990 **53:** 1-6

15 Arrelöv B Alexanderson K, Hagberg J, Löfgren A, Nilsson G, Ponzer S. **Dealing with sickness certification - a survey of problems and strategies among general practitioners and orthopaedic surgeons**. *BMC Public Health* 2007 **7**: 273

16 Swartling M Peterson S, Wahlström R.**Views on sick-listing practice among Swedish General Practitioners--a phenomenographic study**. *BMC Fam Pract*. 2007 **8**: 44

17 Peric F. **The general practitioner's role in the Swiss Invalidity Insurance**. *Schweiz Rundsch Med Prax*. 2005 **94**: 1345-7

18 Toulmin SE. **The use of arguments**. (Updated version) Cambridge Cambridge University Press 2003 p 92

19 Steenbeek R, Boer WEL **de. Mediprudence phase 2: 8 case descriptions** *[in Dutch]* Hoofddorp TNO 2006

20 Health council of the Netherlands **Advice on mediprudence** *[in Dutch]* 2007/14 Den Haag Gezondheidsraad 2007

21 Krohne K, Brage S. **New rules meet established sickness certification practice: a focus-group study on the introduction of functional assessments in Norwegian primary care**. *Scandinavian Journal of Primary Health Care* 2007 **25**: 172-7

22 Governmental committee on working incapacity. **Working on working capacity** *[in Dutch]* Den Haag Ministerie SZW 2001
23 Prinz C. **European Disability Pension Policies**. 11 Country Trends 1970-2002. Vienna European Centre 2003
24 Waddell G., Aylward M., Sawney Ph. **Back pain, incapacity for work and social security benefits. An international literature review and analysis**. London Royal Society of Medicine Press 2003
25 Morgan, D. **Focus groups as qualitative research**. London Sage 1988
26 Kitzinger, J. **The methodology of focus groups: the importance of interaction between research participants**. In: Nigel Fielding (ed.). *Interviewing volume 1*. Sage Benchmarks in social research methods. Pp 347-364 London Sage 2003
27 Boer WEL de, Steenbeek R. **Pilot study on mediprudence: four case descriptions** *[in Dutch]* Hoofddorp TNO 2005
28 World Health Organisation (WHO). **International Classification of Functioning, Disability and Health (ICF)**. Geneva WHO 2001
29 Slebus FG Sluiter JK, Kuijer PP, Willems JH, Frings-Dresen MH.**Work-ability evaluation: a piece of cake or a hard nut to crack?** Disab and Rehab 2007 **29**: 1295-300

Appendix:

Aspects of disability, grounds, and arguments by country.

Grounds	Arguments	Country
Aspect 1. Grounds on claimant's health condition		
10 It is possible that a health condition is so severe that any form of work is excluded	10.1 Clinical and functional impairments are not too severe to prevent him from doing any kind of work	Be, Si
11 It is possible that a health condition is severe to an extent that it precludes from some work but not all work	11.1 His level of functioning is low, too reduced for partial disability	No
12 Disability represents a restriction of functional capacities	12.1 Medication (tramadol) causes a lack of alertness	Nl, Si
	12.2 Several work activities (heavy lifting and carrying etc) are well known risk factors for low back pain and should be avoided	Nl, Si
	12.3 Physical examination, complaints, and functioning are consistent with each other	Nl
	12.4 This kind of chronic pain is tiring and leads to restrictions in energetic activities	Nl
	12.5 Claimant may over esteem his capacities	Be
	12.6 Complaints and claimed restrictions are severe	Be
	12.7 There are no other explanations for his reduced functioning but disease and possibly a weak motivation	No
	12.8 There are no signs of a lack of control of his functioning or alertness	Nl
	12.9 Clinical and functional impairments are not too severe to prevent him from doing any kind of work	Be, Si
	12.10 His reduced functioning remains not fully explained with the medical findings	No
	12.11 His level of functioning is low, too reduced for partial disability	No
13 Capacity for work represents the ability to perform jobs	13.1 An unqualified worker can be referred to many jobs, including light work	Be

	13.2 He has the possibility to adapt himself to other work	Be
14 Advanced age can be a reason to accept restrictions in activities.	14.1 His age is not a big problem	Be
Aspect 2. Grounds on a proper process of evaluation		
20 Findings have to be plausible	20.1 He seems well motivated to work and follows medical advice	Nl
	20.2 His medication causes a lack of alertness. He notices this effect himself	Nl
	20.3 There is pathologic evidence of damage of back and knee	Be, Nl, Si, No
	20.4 Several work activities (heavy lifting and carrying etc) are well known risk factors and should be avoided	Si, Nl
	20.5 Trial of work resumption failed	No
21 Findings have to be consistent	21.1 Examination is consistent with some incapacity	Be
	21.2 Physical examination, complaints and functioning are consistent with each other	Nl
	21.3 His reduced functioning remains not fully explained with the medical findings	No
	21.4 There are inconsistencies between complaints and findings	Nl,
	21.5 Claims to need strong medication	Be
	21.6 He does not complain about fatigue with his pain	Nl
	21.7 He has been properly examined	No
22 Restriction of abilities must not be explained by other factors, notably lack of motivation or opportunity to function	22.1 He might comply better with medical advice, is obese, has a pessimistic view on his future and is inactive	Nl
	22.2 He seems well motivated to work and follows medical advice	Nl
23 In order to determine a claimant's abilities his personal experience is a source	23.1 His medication causes a lack of alertness. He notices this effect himself	Nl

	23.2 Claimant may over esteem his capacities	Be
	23.3 Complaints and claimed restrictions are severe	Be
	23.4 He claims to be unable to sustain efforts with his back	
	Nl	
24 In order to determine a claimant's abilities the medical diagnosis is a source	24.1 He has pathological degeneration of his back and knees. This risks further damage. Several work activities (heavy lifting and carrying etc) are well known risk factors and should be avoided	Nl, Si, Be,
	24.2 Many people with this condition do work	Be
25 In order to determine a claimant's abilities the medication is a source	24.3 His medication causes a lack of alertness. He notices this effect himself	Nl, Si,
Aspect 3. Grounds on treatment, rehabilitation, and time perspective		
30 Disability can be accepted as permanent when all treatment options have been tried	30.1 He has had all treatment necessary	No
	30.2 He has had all rehabilitation possible	No
	30.3 Trial of work resumption failed.	No
	30.4 It is really time he should get back to a job	Be
	30.5 Treating neurologist thinks further treatment useless and restrictions severe	Nl
Aspect 4. Grounds on efforts to recover and resume work		
40 If possibilities for treatment, rehab and/ or work resumption exist the claimant is requested to try these	40.1 He could use a different medication.	Nl
	40.2 He might comply better with medical advice, is obese, has a pessimistic view on his future and is inactive	Nl
	40.3 Maybe treatment options still exist	No
	40.4 It is really time he should get back to a job	Be
	40.5 Further treatment is possible, like graded activity	Nl

5 Grounds of medical evidence		
51 Tramadol (an opiate) can cause a lack of alertness	51.1 His medication causes a lack of alertness. He notices this effect himself	Be, Nl, Si
52 Heavy lifting, carrying and the like are well known risk factors for low back pain and should be avoided in the work of people who suffer from low back pain	52.1 Several work activities (heavy lifting and carrying etc) are well known risk factors for low back pain and should be avoided	
52.2 He has pathological degeneration of his back and knees. This risks further damage. Several work activities (heavy lifting and carrying etc) are well known risk factors and should be avoided	Be, Nl, Si	
53 Chronic low back pain is tiring and may lead to restriction of energetic activities	53.1 This kind of chronic pain is tiring and leads to restrictions in energetic activities	Be, Nl, Si, No
54 Pathologic damage of back and knee make complaints of back and knee plausible	54.1 There is pathologic evidence of damage of back and knee	Be, Nl, Si, No

Chapter 4:
Guidelines in assessment of work disability: an international survey

WEL de Boer[1], AM Rijkenberg[2], P Donceel[3]

[1] Department of Quality of Life, TNO, Hoofddorp, the Netherlands
[2] At the time: Department of Quality of Life, TNO, Hoofddorp, the Netherlands. At present: Occup. Med. Service, OCMW-Antwerpen, Antwerp, Belgium
[3] Section of Occupational, Environmental and Insurance Medicine, Department of Public Health, Katholieke Universiteit Leuven, Leuven, Belgium

Abstract

Background
Assessments of work disability are carried out by social insurance physicians (SIPs) and little supported with evidence or instruments. Guidelines are hardly used in social insurance medicine. Development in social insurance medicine might be slow as insurance is different from clinical medicine.

Aims
We explored comparability of assessments in social insurance medicine in different countries and what guidelines were in official use.

Methods
Eighteen European countries were invited. A questionnaire on assessments practices was sent to national experts. A comparative table was presented to all contributors. Countries with guidelines were visited. Guidelines were categorised according to their prupose in their content was compared. The results were proposed for validation to experts of participating countries.

Results
Fourteen countries participated. Functional capacity assessment was common. Guidelines for SIPs were reported to be officially in use in Germany, Ireland, the Netherlands and Switzerland. Twenty-two guidelines were medical and eleven were procedural. Medical guidelines mainly treat the same topics. Procedural guidelines are more variable.
Assessment of work disability is comparable between countries. Medical and procedural guidelines can be further developed and tested on their value in practice. The procedural guidelines need to be published in a clear and comparable manner. The legal security of claimants would be endorsed by this. Germany and the Netherlands are most experienced and could take the lead in international development.

Introduction

In the western world, work disability is a problem at both the individual level and the level of the company and society. In many compensation systems, social insurance physicians (SIPs) assess and legitimise the claim for entitlement to long-term disability benefits [1]. These disability assessments are traditionally based mainly on legislation, rules of the administration, and doctors' expertise. The quality of these assessments is not easy to establish as there is no gold standard for their quality [2, 3, 4]. One way of controlling the quality of medical work is to use guidelines [5], which is common in clinical practice. The medical practice in social insurance medicine is different from clinical medical practice in several ways [6]. In clinical practice the consultation is a private initiative of a patient who seeks help that is often restricted by policies of health insurance, whereas in social insurance the consultation is an assessment determined by the legal context and the constraints of the implementing body, the Institution of Social Insurance (ISI). In clinical practice the focus is on disease and curing, whereas in social insurance medicine the focus is on problems with capacity to work and return to work. In clinical practice the patient's request for treatment is taken for granted, whereas in social insurance the claim to be exempt from work and for a benefit is scrutinized and evaluated. These differences have been found to influence the practice of the assessments [7].

Guidelines for assessing disability for work in social insurance are a recent development. Guidelines were first created in the Netherlands in 1996 and were developed in reaction to changes in organisation and legislation. A second wave of guidelines occurred from 2005 on, in reaction to upcoming changes in legislation. The guidelines are meant to support the SIP but also to enhance transparency of the evaluations and the legal position of the client. For this the guidelines need to be officially published and accessible to clients. There is little literature with regard to development of guidelines for assessing disability in social insurance. We were interested to know what guidelines existed in different countries.

Schemes for disability are known to vary in criteria and organisation among countries [8, 9]. For example, in compensation schemes for work accidents, physical damage is often the topic of assessment, whereas in general disability schemes, assessment of functional capacity is more common. In terms of ICF [10] this is a difference of level: impairments versus activity restrictions. This might hamper comparison of medical guidelines used for these different purposes. These differences and similarities in evaluation of work disability among countries have barely been addressed in the literature. It is reported, however, that the medical assessments in social insurance concentrate on similar issues. Brage et. al [11] found it possible to make a core set of ICF items that could be used internationally. Boer et. al. [12] found, in four countries, the reasoning of the physicians that underlies their conclusions on disability to represent the concept of the

handicapped role [13] or disability role [6]. These SIPs also strived to establish a fair trial.
The concept of the handicapped role describes the rights and obligations of people with disabilities towards society. In short, this entails the right to be (partly) exempt from work and to receive care, and the obligation to strive for recovery as much as possible and to take up work as soon as possible, and finally, to account for all of this. In social insurance the accounting occurs via SIPs who evaluate disease, functional capacity, prognosis and possibilities for therapy, rehabilitation, and return to work activities. The concept of a fair trial refers to the right of the claimant to receive a proper assessment by a neutral and competent professional who allows the claimant to state his claim and his arguments. If these concepts are indeed central to medical assessments in social insurance in different countries, this means that medical practice in social insurance may be comparable among countries. The quality of the assessments can be supported with guidelines for assessment that may be comparable as well.

The research questions of this study are:

1 Do medical assessments of work disability in different countries address the handicapped role?
2 What kinds of guidelines that support the assessment of work disability are officially published?

Methods

1 European countries were approached through EUMASS, a network of associations of social insurance medicine in European countries (www.eumass.com). This was done for practical reasons so as to be able to reach many countries. All seventeen members were invited through the council members, who usually are the central medical advisers in social insurance. Switzerland, not a EUMASS-member, was also invited, as it had developed at least one guideline. We developed a questionnaire about how the assessments were performed, indicating the key aspects of the medical assessment that were formulated in terms of the handicapped role. To promote uniform data collection, the questionnaires contained examples of the Dutch answers. The questionnaires were sent to experts that were indicated by the representatives of the countries in the EUMASS council. These contact persons were free to consult other experts. When needed, clarification was provided through email or telephone contact. The final comparative table was presented to all contributors.

2 These questionnaires asked about the use of instruments to support the assessment of disability for work. Countries in which guidelines were reported to be in use (Czech Republic, Germany, the Netherlands, the

United Kingdom, and Switzerland) were visited. Ireland was not visited but explored through personal and mail contact. The guidelines were explained and their status was discussed and a copy was collected during the visits. For this article we focused on the guidelines for assessments of incapacity for work by SIPs that were in official use; that is, prescribed by law or as formal instruction by the administration of the ISI. We found two types of guidelines: medical and procedural. Medical guidelines describe aspects of assessment in cases of a specific pathology, which is clearly identified in the title of the guideline. Procedural guidelines describe aspects of assessment regardless of pathology, which, again, is identified in the title of the guideline. To compare the medical guidelines we (WB + AR) compared the diagnoses that were subject of the guidelines and we compared the structure of the guidelines. We used the handicapped role as a starting point for comparing the structure, adding topics that were treated. To compare the procedural guidelines, we drafted a format of aspects of fair trial and scored which topic was treated in which guideline. We used the topics of the Dutch guidelines as a starting point, as we were most familiar with these and as these were the most detailed. When initial scoring did not lead to agreement, a consensus discussion was held and, if necessary, PD would arbitrate. The results were proposed for validation to experts of participating countries.
Ethical approval was not necessary for this study.

Results

1 Of seventeen countries associated in EUMASS, thirteen participated until the final result; Switzerland also participated. The core aspects of the medical assessments were structured using the concept of the handicapped role: disease, functional capacity, prognosis, treatment/ rehabilitation, and return to work. The wording of the aspects claimants are assessed on differs among countries. Disease is sometimes mentioned as illness or health condition or health problem. Functional capacity is sometimes called working capacity or earning capacity. Treatment and rehabilitation may be referred to as activities to promote return to work. Appendix A presents the precise wording. The results are shown in Table 1.

Table 1 assessment aspect of handicapped role according to country

Aspect	Country
Disease	Be, Ch, De, Fi, Fr, Ie, It, Nl, No, Si, Sl,
Functional capacity	Be, Ch, Cz, De, Fi, Fr, Ie, It, Hu, Nl, No, Se, Si, Sl,
Treatment/ rehab	Ch, Cz, De, Fi, Nl, No,
Prognosis	Be, Ch, Cz, De, Fi, Ie, It, Hu, Nl, Sl,

In all countries functional capacity was assessed. In eleven countries disease was an aspect of the assessment, in ten countries the prognosis, and in six treatment and rehabilitation were assessed.

2 Guidelines for SIPs were reported to be officially in use in Germany (7), Ireland (1), the Netherlands (24), and Switzerland (1). Twenty-two guidelines were medical and eleven were procedural. This distinction was evident in the titles of the guidelines. The guidelines are given in Appendix B.

Six medical guidelines, published by the German Institute of Social Insurance (DRV), were in official use in Germany as administrative prescriptions. In the Netherlands sixteen were used as legal prescriptions. These were partly published by the National Health Council and partly by the Dutch Association of Social Insurance (NVVG). In the Czech Republic, a barema type of guideline is in official use, which was excluded from this study as it does not assess functional capacity but impairments. The remaining guidelines all refer to a specific pathology and are shown in Table 2.

Table 2: Medical guideline, year of publication / revision, reference number

Guideline NI	Guideline De
Non Specific Lumbar Disorder 2005/2008	
Myocardial Infarction 2005/ 2008	Coronary Heart Disease 2001/ 2005
Chronic Heart Failure 2007	
Anxiety Disorders 2007	
Stroke 2007	
Breast cancer 2007	Breast Cancer 2006
Chronic Fatigue Syndrome 2007	
Herniating intervertebral disc 2007	Herniating intervertebral disc 2002/ 2003/ 2005
Burn out 2007	
Depressive Disorder 2007	
Whiplash Associated Disorders 2008	
Arthritis Hip and Knee 2007	
Rheumatoid Arthritis 2007	
Chronic Obstructive Lung Disease 2007	Chronic Obstructive Lung Disease 2003/ 2005
Schizophrenia and associated disorders 2007	
Chronic Shoulder Disorders 2007	
	Mental disorders 2001/ 2006
	Chronic Inflammatory Bowel Disease 2005

The Dutch medical guidelines, which were first developed by the Health Council of the Netherlands and later by the scientific association of SIPs (NVVG), are all implemented by law. The German guidelines were developed by the German Institution of Social Insurance (DRV) and all are administrative prescriptions. In Germany, the guidelines have been revised several times, which is not the case in the Netherlands.

The structure of the guidelines was compared starting from the structure of the Dutch guidelines. WB and AR reached consensus on the following: All Dutch guidelines have the same structure as indicated by the contents page, as do the German guidelines. This is presented in Table 3.

Table 3: Structure of Dutch and German guidelines according to the contents page

Dutch	German
Guideline itself (summary of recommendations)	
Scope and purpose, method of development	Scope and purpose, method of development
The medical condition itself	The medical condition itself
Origin, risk factors, course	Risk factors, course
Diagnostic procedures, treatment, ICD classification	Diagnostic procedures, ICD classification
Return to work activities	
Assessment of disability for work	Assessment of functional capacity
Literature	Literature
Involved experts and organisations	Involved experts and organisations
	Annexes (tools for assessment)

Both types of guideline are very similar, treating mainly the same topics. The Dutch guidelines have a separate section with recommendations and a section devoted to return to work activities. The German guidelines have several tools for assessment in their annexes.

Procedural guidelines were in use in Germany: "Medical assessment for the disability insurance" 77p, 2001; Ireland: "Medical assessments," 2 p no year; the Netherlands: "Medical criterion of incapacity for work," 22p 1996, "Permanent full disability,"25p, "Communication with treating physicians," 17 1996, "Assessment methods," 16p 2000, "Report protocol social insurance medicine," 9p 1999, "Restriction of working hours," 25p 2000, "Professional criteria for reassessment," 7p 2000, "Increased disability by the same cause," 3p, no year, and "Code of conduct"; and Switzerland: "Assessment in insurance medicine," 2007. The German, Irish, and Swiss guidelines each contain recommendations that in the Netherlands are distributed into eight smaller guidelines. The Irish recommendations are brief and general, while the other countries provide more detailed recommendations and rationale

for their validity. The recommendations in Germany, the Netherlands, and Switzerland represent consensus of medical and often legal experts but not scientific evidence. We listed thirteen topics of the Dutch guidelines and compared the other guidelines to these. The Swiss guideline contained two more topics, which we reached consensus about. Experts in all four countries commented on the results and agreed with them. The results are presented in Table 4.

Table 4 topics of procedural guidelines to country

Topic	CH page	DE page	IE paragr.	NL title of guideline
Concepts of disease and (in) capacity to work/ earn as assessed in social insurance	33, 50-56	6, 7, 13-15	Par 5	medical criterion of incapacity for work
Independence of SIP between claimant and ISI	10, 11	16-17	Par 4, 5	medical criterion of incapacity for work code of conduct
Qualification of SIP	14	16-18	Par 4	medical criterion of incapacity for work
Methods to determine the existence of disease	11, 13, 14	6-11 17-20	Par 5	Assessment methods
Methods to determine functional capacity	14-16	17-20, 39-44, 55-57	Par 5	Assessment methods
Methods to calculate earning capacity				Cooperation SIP labour expert
Quality criteria of assessments	23-35	21-28, 58-60		report protocol SIP
General criteria for full incapacity for work		13-14		Permanent full disability
General criteria for permanent incapacity	56-58	6, 7, 28, 67		Permanent full disability
General criteria for restriction of functional capacities		13,27, 66-67		restriction of working hours
General criteria to contact the treating physician		18, 44, 45		communication treating physicians
Procedure to contact the treating physician		18, 44, 45		communication treating physicians
Protection of private information	24-31	18		management of medical information
Medical secret		18		management of medical information.
Obligation to mitigate the damage	58-60	10		
Situations where increase of disability is due to the same cause as earlier disability				increased disability by the same cause
Professional criteria to reassess a client who receives a benefit		43		Professional criteria for reassessment

Experts of Germany, the Netherlands, and Switzerland commented that the distinction between legal and administrative instructions was fuzzy. In other legal texts instructions are given that also influence the assessments.

Discussion

In this study we wanted to address the following questions:

1 Do medical assessments of work disability in different countries address the handicapped role?
2 What kinds of guidelines to support the assessment of work disability are officially published?

Main findings

Using the EUMASS network and Switzerland, we found that the aspect of assessing functional capacity was common to all fourteen responding countries. In terms of ICF this refers to activity limitations. Treatment and rehabilitation were reported to be assessed only in six countries. Guidelines, both procedural and medical, were in official use in four countries. In two countries we found twenty-two medical guidelines in official use. Four pathologies were common, eighteen were specific to countries. Topics are largely similar, with the Dutch guideline paying specific attention to return to work aspects and the German guidelines providing tools for assessment. In four countries we found eleven procedural guidelines, which differed greatly in length, from two to seventy-seven pages. Five topics were common to all four countries, one to three countries. In countries that have both, procedural guidelines were developed before medical guidelines. We found no country with medical guidelines and without procedural guidelines, but we did find the opposite in Ireland and Switzerland.

Strengths and weaknesses

To our knowledge, this is the first study that aims at identifying and qualifying guidelines in social insurance medicine at an international level. Using the EUMASS network and Switzerland we were able to address the central medical advisers who are in a good position to answer questions on guidelines, as they are responsible for the medical processes in their institutions. Using two rounds of questionnaires and a check on all results by all participants, we believe the responses have good reliability. Still, some questions may not be resolved.

The question of what claimants are assessed on may be answered incompletely in our study when it comes to daily practice. In several countries the aspect of disease was not mentioned, but it is unlikely that it plays no role. This applies equally for the aspect of treatment and rehabilitation. If we had observed SIPs in practice, we might have found even more similarities For our results, however, which focus on official guidelines for the assessments, what SIPs do in daily practice probably makes no difference.

Our focus on official guidelines resulted in finding fewer guidelines than

are in practical use. For example, in Germany and Switzerland, guidelines are published by specialists in scientific journals. These are not considered official, but may impact practice. In the UK, medical guidance is provided for the disability analyst who is not a medical doctor. The distinction between legal guidelines and legal and administrative procedural instructions is fuzzy, so it is likely that more instructions exist than we found. Indeed, it is very likely that aspects such as the neutral position of the assessor, the concept of disease in social insurance, etc., have crystallised into some form of instruction everywhere. Interestingly, these have not been brought up by our respondents on our question about guidelines, possibly because the respondents did not associate procedural issues with the concept of guidelines. A separate study would be needed to elicit the rules of the game of the assessments, be they official guidelines or something else.

We listed the guidelines as comparable to each other. For the medical guidelines this is probably sufficient, although a comparison of their content is needed to be certain. For the procedural guidelines this is doubtful. They are all officially prescribed, but this will have a different impact as they are so different in extensiveness. Here too a content analysis is needed to establish their comparability.

The handicapped role seems to be a reliable basic concept to study these assessments; however, this may not be true for the concept of a fair trial in the assessments. The way in which we operationalised it will need further testing, as we restricted our study to guidelines and their topics. Another matter is the quality of the recommendations in the guidelines, which we did not address.

Other studies
Our study corresponds to other research. The accent on assessing functional capacity is the modern approach in social insurance medicine [2, 14]. The distinction between procedural and medical guidelines fits the results of other research [12, 15] about the medical aspect of the doctor's reasoning, in combination with the procedural aspect.

Impact
Our results are restricted to existing guidelines in four countries. We expect our results to be relevant guidelines that are to be developed in the same or other countries. The context of social insurance calls for two types of guidelines to ensure quality of assessment: medical and procedural. The history of development suggests that agreement on procedure precedes agreement on medical content.

In many countries the quality of the assessments cannot be explicitly tested with established guidelines. This seems unsatisfactory, both from a professional perspective and from a perspective of legal security of the

claimant. Good practice is best supported by guidelines that have been developed in a proven optimal way. The medical guidelines support the assessment of disability by proving established grounds for what can be expected with certain pathologies. Medical guidelines can be tested with the AGREE instrument [16, 17]. The procedural guidelines support the use of a fair trial when evaluating an individual's right to financial compensation. For procedural guidelines, no instrument exists to establish their quality. Using the results of our study, a general format for fair trial could be developed and tested in different countries. The procedural guidelines could be linked to official documents such as the Convention for the Protection of Human Rights and Fundamental Freedoms [18] so as to ground them on universal legal principles. Doing so would make this part of the assessment considerably more transparent. A further step would be to compare the recommendations that are presented in the various guidelines. Reaching agreement on valid recommendations would considerably contribute to the profession of social insurance medicine. However, to achieve this, the guidelines need to be available in English.

Conclusion

We found the assessment of work disability to be comparable among countries, using the handicapped role as central concept. Official guidelines in social insurance medicine, for assessment of work disability, are restricted to Germany, Ireland, the Netherlands, and Switzerland. Guidelines with a medical focus seem well developed; procedural guidelines are much more divergent. Both medical and procedural guidelines can be further developed and tested on their value in practice. For this they should be officially published, preferably also in English, and scientifically tested. For international comparison the wording should be much more precise and operationalised. ICF and the handicapped role are useful in harmonising the language. Germany and the Netherlands could take the lead in this as these countries are the most experienced. For the practice of assessment of work disability, well-founded guidelines would be important instruments to control quality. The legal security of claimants and the profession of social insurance medicine would be endorsed by this.

Authors' contribution: Wout E.L. de Boer designed the study, carried out the field work, and prepared the manuscript. Arie M. Rijkenberg participated in the field work. Peter Donceel supervised the field work and participated in drafting the article.

Acknowledgements: The study was supported with a grant from the SIG Foundation, which had no involvement in the study itself or in the decision to submit the paper for publication.

Competing interests: All author(s) declare that they have no competing interests.

References.

1 Boer WEL de, Besseling JJM, Willems JHBM. **Organisation of disability evaluation in 15 countries.** *POS* 2007 **38**: 205-217

2 Wind H. **Assessment of physical work ability: the utility of functional capacity evaluation for insurance physicians.** Ph D thesis. Amsterdam Amsterdam University. 2007

3 Überschär 2008 **Quality assurance in the socio-medical assessment in the German Pension Insurance**.*[in German] Gesundheitswesen* 2008; **70**: 690-695

4 Verbeek J Dijk FJH van. **Assessing the ability to work**. BMJ 2008 **336**: 519-520

5 Burgers JS, Cluzeau FA, Hanna SE, Hunt C, Grol R. **Characteristics of high-quality guidelines: evaluation of 86 clinical guidelines developed in ten European countries and Canada**. *Int J Technol Assess Health Care* 2003 **19**: 148-57

6 Waddell G, Aylward M. **The scientific and conceptual basis of incapacity benefits**. Norwich TSO 2005

7 Boer WEL de, Wind H, Dijk FJH van, Willems JHBM. **Interviews for the assessment of long-term incapacity for work: a study on adherence to protocols and principles**. *BMC public health* 2009 **9**: 169

8 Council of Europe. **Assessing disability in Europe**. Strasbourg; CoE 2002

9 Mabbet D, Bolderson H, Hvinden B. **Definitions of disability in Europe: a comparative analysis**. Uxbridge Brunel University 2002

10 World Health Organisation (WHO). **International Classification of Functioning, Disability and Health (ICF)**. Geneva WHO 2001

11 Brage S, Donceel P, Falez F. **Development of ICF core set for disability evaluation in social security**. *Disab & Rehab* 2008 **30**: 1392-1396

12 Boer WEL de. Brage S. Donceel P. Rus M. Willems JHBM. **Medicolegal reasoning**. *BMC Public Health* 2008 **8**: 335

13 Gordon G **Role theory and illness**. New Haven College and university press 1966

14 Organisation for Economic Co-operation and Development (OECD). **Transforming Disability into Ability. Policies to Promote Work and Income Security for Disabled People**. Paris OECD 2003

15 Berg M, Horstman K, Plass S, Heusden M van. **Guidelines, professionals and the production of objectivity: standardisation and the professionalism of insurance medicine**. *Soc Hlth Illness* 2001 **22**: 765-791

16 AGREE collaboration. **Appraisal of Guidelines for Research & Evaluation**. www.agreecollaboration.org 2001

17 Boer WEL de, Bruinvels DJ, Rijkenberg AM, Donceel P, Anema JR. **Evidence-based guidelines in the assessment of disability for work: an international survey and a comparison of quality of development according to AGREE in Germany and the Netherlands**. *BMC public health* 2009 **9**: 349

18 **Convention for the Protection of Human Rights and Fundamental Freedoms** Council of Europe Strasbourg CoE 1950

Appendix A:

Assessment of long term incapacity for work in European countries

Country	Belgium	Czech Republic	Finland (Tyel) Private scheme
Aspects the claimants are assessed on.	- New or aggravating health problem that causes work cessation - Loss of earning capacity of over 2/3 with regard to suitable reference jobs. - Duration of benefit	- Reduced ability to participate in gainful activity in percentage (at least 33 % for partial disability, at least 66 % for full disability). - possibility of RTW actions - Permanence of this incapacity.	- Illness, defect or injury that causes a decrease in earning or prevents claimant from working for over 40 % (or over 60 % partial disability) - Treatment and - Vocational / Non-vocational rehabilitation. - Prognosis of at least a year of incapacity.

Country : 	Finland (Kela) Public scheme	France	Germany
Aspects the claimants are assessed on.	- Is it an illness, defect or injury that causes a decrease in earning or prevents claimant from working. - Possibility of treatment and rehabilitation first - Prognosis of permanent incapacity.	- Health condition - Loss of functional capacity of 2/3 or over. - Loss of earning capacity of 2/3 or over	Rehabilitation (medical and RTW) is investigated. Prognosis Aetiology Work incapacity in working hours (Complete: 0 – 3 hours / Partial: 3 – 6 hours / None: 6 – 8 hours), What disabilities does the client have. Work intensity, Work organisation.

Country	Hungary	Ireland	Italy
Aspects the claimants are assessed on.	functional capacity earning capacity prognosis	Health condition capacity for all types of work permanence of this incapacity	- Health condition (afflicted by a mental or physical infirmity). - Degree of restriction or loss of working capacity of or over 2/3 in any of the activities one had an aptitude for, or absolutely and permanently unable to perform any gainful activity - Prognosis (disability must be permanent).

Country	The Netherlands	Norway	Slovakia
Aspects the claimants are assessed on.	- Disease or handicap that reduce - Functional capacity in max. 56 items that causes - Loss of earning capacity between 35% and 80% or over 80 % - Rehabilitation and RTW activities in preceding two years of sick leave - Actual possibilities for treatment and RTW - Prognosis	- Permanent disease, injury or impairment as cause of the loss of function that is the main cause of the loss of over 50% of earning capacity - Adequate treatment and vocational rehabilitation with regard to age, capacities, education and possibilities for work. - Effort to RTW	Health condition Prognosis Notably impairments (in percentage from 41 % upwards) Loss of earning capacity (Between 41 – 70 % and 71 upwards).

Country	Slovenia	Sweden	Switzerland
Aspects the claimants are assessed on.	Change of health condition that reduces claimant's possibilities for getting or keeping one's job or getting promoted in one's career. Three categories of invalidity.	Loss of earning capacity (25%, 50%, 75%, 100%) Prognosis	Health condition Capacity Prognosis Rehabilitation

Appendix B:

officially prescribed guidelines for evaluation of work disability in Germany, Ireland, the Netherlands and Switzerland

Germany:
Deutsche Rentenversicherung **Leitlinien zur sozialmedizinischen Leistungs-beurteilung bei koronarer Herzkrankheit** [guideline for the socio-medical assessment of coronary heart disease] 2005 www.drv.de Deutsche Rentenversicherung **Leitlinien zur sozialmedizinischen Beurteilung der Leistungsfähigkeit bei Mama-Karzinom** [guideline for the socio-medical assessment of Breast Cancer] 2006 www.drv.de
Deutsche Rentenversicherung **Leitlinien zur sozialmedizinischen Beurteilung der Leistungsfähigkeit bei Bandscheiden und Bandscheibenassoziierten Erkrankungen** [guideline for the socio-medical assessment of Herniating intervertebral disc] 2005 www.drv.de
Deutsche Rentenversicherung **Leitlinien zur sozialmedizinischen LeistungsbeurteilungbeichronischobstruktivenLungenkrankheiten (COPD) und Asthma Bronchiale** [guideline for the socio-medical assessment of Chronic Obstructive Pulmonary Disease and Bronchial Asthma] 2005 www.drv.de
Deutsche Rentenversicherung **Leitlinien für die sozialmedizinische Beurteilung von Menschen mit psychischen Störungen** [guideline for the socio-medical assessment of people with mental disorders] 2006 www.drv.de
Deutsche Rentenversicherung **Leitlinien zur sozialmedizinischen Leistungsbeurteilung bei chronisch entzündlichen Darmkrankheiten** [guideline for the socio-medical assessment of chronic inflammatory bowel disease] 2005 www.drv.de
Deutsche Rentenversicherung **Das ärztliche Gutachten für die gesetzliche Rentenversicherung; Hinweise zur Begutachtung** [Medical guideline for the assessment], 2001 www.drv.de

Ireland:
Medical assessments no year www.welfare.ie/

The Netherlands
Gezondheidsraad **Verzekeringsgeneeskundig protocol Angststoornissen** [Guideline for social insurance physicians: Anxiety Disorders] 2007 www.gr.nl 2007/05
Gezondheidsraad **Verzekeringsgeneeskundig protocol HNP** [Guideline for social insurance physicians: Herniating intervertebral disc] 2007 www.gr.nl 2007/12
Gezondheidsraad **Verzekeringsgeneeskundig protocol Depressie** [Guideline for social insurance physicians: Depressive Disorder] 2007 www.gr.nl 2007/22

Gezondheidsraad **Verzekeringsgeneeskundig protocol Borstkanker** [Guideline for social insurance physicians: Breast cancer] 2007 www.gr.nl 2007/05
Gezondheidsraad **Verzekeringsgeneeskundig protocol Chronisch Vermoeidheidssyndroom** [Guideline for social insurance physicians: Chronic Fatigue Syndrome] 2007 www.gr.nl 2007/12
Gezondheidsraad **Verzekeringsgeneeskundig protocol Burn out** [Guideline for social insurance physicians: Burn out] 2006 www.gr.nl 2006/22
Gezondheidsraad **Verzekeringsgeneeskundig protocol CVA** [Guideline for social insurance physicians: Stroke] 2007 www.gr.nl 2007/05
Gezondheidsraad **Verzekeringsgeneeskundig protocol Whiplash Associated Disorders** [Guideline for social insurance physicians: whiplash associated disorders] 2008 www.gr.nl 2008/11
Gezondheidsraad **Verzekeringsgeneeskundig protocol Aspecifieke rugklachten** [Guideline for social insurance physicians: Non Specific Lumbar Disorder for Social Insurance Physicians] 2008 www.gr.nl 2008/11
Gezondheidsraad **Verzekeringsgeneeskundig protocol Hartinfarct** [Guideline for social insurance physicians: Myocardial Infarction] 2008 www.gr.nl 2008/11
NVVG **Verzekeringsgeneeskundig protocol Chronisch Hartfalen** [Guideline for social insurance physicians: Chronic Heart Failure] 2007 www.nvvg.nl
NVVG **Verzekeringsgeneeskundig protocol arthrose van heup en knie** [Arthritis Hip and Knee] 2007 www.nvvg.nl
NVVG **Verzekeringsgeneeskundig protocol Reumatoide Artritis** [Guideline for social insurance physicians: Rheumatoid Arthritis] 2007 www.nvvg.nl
NVVG **Verzekeringsgeneeskundig protocol COPD** [Guideline for social insurance physicians: Chronic Obstructive Lung Disease] 2007 www.nvvg.nl
NVVG **Verzekeringsgeneeskundig protocol Schizofrenie en aanverwante stoornissen** [Guideline for social insurance physicians: Schizophrenia and associated disorders] 2007 www.nvvg.nl
NVVG **Verzekeringsgeneeskundig protocol chronische schouderklachten** [Guideline for social insurance physicians: Chronic Shoulder Disorders] 2007 www.nvvg.nl
NVVG Richtlijn **Medisch Arbeidsongeschiktheidscriterium** [Guideline for social insurance physicians: medical criterion of incapacity for work], 1996 www.nvvg.nl
NVVG Standaard **Geen duurzaam benutbare mogelijkheden** [Guideline for social insurance physicians: permanent full disability] 1996 www.nvvg.nl
NVVG **Standaard Communicatie met behandelaars** [Guideline for social insurance physicians: communication with treating physicians] 1996 www.nvvg.nl

NVVG **Standaard onderzoeksmethoden** [Guideline for social insurance physicians: assessment methods] 2000 www.nvvg.nl
NVVG **Rapportageprotocol** [Guideline for social insurance physicians: report protocol social insurance medicine] 1999 www.nvvg.nl
NVVG **Standaard urenbeperking** [Guideline for social insurance physicians: restriction of working hours] 2000, www.nvvg.nl
NVVG **Standaard professionele herbeoordeling** [Guideline for social insurance physicians: professional criteria for reassessment] 2000 www.nvvg.nl
NVVG **Standaard toegenomen arbeidsongeschiktheid door dezelfde oorzaak** [Guideline for social insurance physicians: increased disability by the same cause] no year www.nvvg.nl
NVVG **Gedragscode verzekeringsarts** [code of conduct Social Insurance Physician] www.nvvg.nl

Switzerland:
Riemer Kafka G.(ed) . **Versicherungsmedizinische Gutachten** [Evaluation in insurance medicine] Stampfli verlag AG Bern 2007

Chapter 5:
Evidence-based guidelines in the evaluation of work disability: an international survey and a comparison of quality of development.

WEL de Boer[1], DJ Bruinvels[2], AM Rijkenberg[3], P Donceel[4] , JR Anema[2]

[1] Department of Quality of Life, TNO, Hoofddorp, the Netherlands.
[2] Department of Public and Occupational Health and EMGO-Institute, Amsterdam, the Netherlands
[3] At the time: Department of Quality of Life, TNO, Hoofddorp, the Netherlands. At present: Occup. Med. Service, OCMW-Antwerpen, Antwerp, Belgium
[4] Section of Occupational, Environmental and Insurance Medicine, Department of Public Health, Katholieke Universiteit Leuven, Leuven, Belgium

Abstract

Background
In social insurance, the evaluation of work disability is becoming stricter as priority is given to the resumption of work, which calls for a guarantee of quality for these evaluations. Evidence-based guidelines have become a major instrument in the quality control of health care, and the quality of these guidelines' development can be assessed using the AGREE instrument. In social insurance medicine, such guidelines are relatively new. We were interested to know what guidelines have been developed to support the medical evaluation of work disability and the quality of these guidelines.

Methods
Five European countries that were reported to use guidelines were approached, using a recent inventory of evaluations of work disability in Europe. We focused on guidelines that are disease-oriented and formally prescribed in social insurance medicine. Using the AGREE instrument, these guidelines were appraised by two researchers. We asked two experts involved in guideline development to indicate if they agreed with our results and to provide explanations for insufficient scores.

Results
We found six German and sixteen Dutch sets of disease-oriented guidelines in official use. The AGREE instrument was applicable, requiring minor adaptations. The appraisers reached consensus on all items. Each guideline scored well on 'scope and purpose' and 'clarity and presentation'. The guidelines scored moderately on 'stakeholder involvement' in the Netherlands, but insufficiently in Germany, due mainly to the limited involvement of patients' representatives in this country. All guidelines had low scores on 'rigour of development', which was due partly to a lack of documentation and of existing evidence. 'Editorial independence' and 'applicability' had low scores in both countries as a result of how the production was organised.

Conclusions
Disease-oriented guidelines in social insurance medicine for the evaluation of work disability are a recent phenomenon, so far restricted to Germany and the Netherlands. The AGREE instrument is suitably applicable to assess the quality of guideline development in social insurance medicine, but some of the scoring rules need to be adapted to the context of social insurance. Existing guidelines do not meet the AGREE criteria to a sufficient level. The way patients' representatives can be involved needs further discussion. The guidelines would profit from more specific recommendations and, for providing evidence, more research is needed on the functional capacity of people with disabilities.

Background

In the western world, work disability is a problem at the individual, company, and societal levels. Western countries spend about 1.2% of GDP on work disability benefits or 2% if sickness benefits are included, which, for most countries, is an increase over the past 15 years. The probability of returning to work after being granted a long-term disability benefit is below 2% annually on average. Work disability is the end of their working life for the vast majority of recipients [1]. To reduce work disability, many countries have restricted access to disability benefits in social insurance and they have developed programmes to promote return to work [2, 3, 4]. In the Netherlands, eligibility criteria have become stricter with the implementation of a new law on long-term work disability. In the United Kingdom, a renewal of the personal capacity assessment for long-term disability benefit was recently implemented [5] and comparable changes are occurring in other countries [2, 3, 4]. These policy changes are meant to result in more people being active in work and fewer people receiving disability benefits. In disability benefit systems, social insurance physicians (SIPs) evaluate claims for entitlement to long-term disability benefits [6]. These work disability evaluations are traditionally based mainly on legislation, administrative rules, and doctors' expertise.

When resources are tight, it becomes even more important to determine in a valid and scientifically sound way who is and is not entitled to disability benefit. Internationally, the medical evaluations of work disability turn out to be relatively comparable while being part of social insurance systems that vary strongly [6, 7, 8]. The quality of these evaluations is not easy to establish, as no gold standard exists for their validity [9, 10]. The mechanism used most often to ensure quality is to organise the process of evaluation in such a way that an optimal result can be expected. A common practice in 14 countries, in Europe and the Russian Federation, is to use qualified doctors, the SIPs, and to have medical reports verified by staff doctors [6]. Although instruments used to support medical decision making are not validated for this purpose [6, 9, 11], this does not necessarily mean that they are unsuitable.

One way of ensuring the quality of medical work is to use evidence-based guidelines [12], which is common in clinical practice [13]. In clinical practice, guidelines, which the clinician can use with his clinical experience and the patient's preferences, are intended to support the physician by providing recommendations for diagnosis, treatment, and prognosis. [14]. Evidence-based clinical practice means using the best evidence available, in consultation with the patient, to decide on the option that suits that patient best [15]. Guidelines, however, are not restricted to clinical practice: some are being introduced on a wider scale in occupational medicine [16, 17] and serve, among other functions, to support the coaching of employees with

work-related health problems [18, 19]. In occupational medicine, guidelines are intended to provide an occupational physician with recommendations for diagnosis and prognosis of the work-related problem and for the selection of effective interventions [17]. These guidelines can be used in addition to the experience of the occupational health professional and the preferences of the employee and employer. However, guidelines for evaluation in social insurance medicine are a rather new phenomenon.

Having guidelines for medical work does not necessarily mean that the quality of the work is supported. Guidelines need to be adequate for the process they are to support and they need to be used in practice. The Appraisal of Guidelines Research & Evaluation (AGREE) collaboration developed the AGREE instrument to assess the quality of clinical practice guidelines [20] and to establish the quality of the development of guidelines with regard to scientific principles. The AGREE instrument is composed of twenty-three items covering six domains of quality of guideline development: 'scope and purpose', 'involvement of stakeholders', ´rigour of development', 'clarity and presentation of recommendations', 'applicability', and 'editorial independence'. The AGREE instrument has been tested in clinical guidelines and was found to have a good reliability [21]. Thus far, there are no universally accepted cut-off points to identify high-quality guidelines [22]. A high-quality guideline can be expected to contribute to high-quality recommendations but does not warrant them as the evidence used is in general limited and controversial [23, 24]. The AGREE instrument is widely used to evaluate clinical guidelines [25, 26], as well as those found in occupational medicine [16, 27, 28], but so far has not been used in social insurance medicine. Social insurance medicine may simply be lagging behind, but the AGREE instrument may not be being used in social insurance medicine because of the rather different medical work involved in social insurance.

Medical practice in social insurance evaluations is different from clinical medical practice in several ways [29, 30]. In clinical practice, the consultation is a private initiative of a patient who seeks help that is often restricted by policies of health insurance, whereas in social insurance medicine the consultation is an evaluation that is determined by the legal context and the constraints that the implementing body, the Institution of Social Insurance (ISI), puts on it. In clinical practice, the focus is on disease and finding a cure, whereas in social insurance medicine the focus is on capacity for, and a return to, work. In clinical practice, a patient's request for treatment is taken for granted; in social insurance medicine, the claim to be exempt from work and for a benefit to be paid is scrutinised and evaluated. The position of the claimant in a social insurance context is therefore different from the position of the patient in a clinical care context, differences that have been found to influence the practice of the evaluations [31]. Furthermore, the position of social insurance physicians is different

from doctors in clinical medicine as the SIPs have an advisory function towards the ISI they work for and not primarily for the claimant [6]. This position may give rise to tensions between administrative procedures for handling big numbers of claimants and the doctors' need to deliver tailor-made evaluations [32, 33].

It is difficult to diagnose the functional consequences of diseases in general and even more so for non-specific diseases such as lower back pain, chronic fatigue, and stress-related disorders. The association between a medical diagnosis and the functional limitations that may lead to work disability is weak and influenced by environmental and personal characteristics, as described in the International Classification of Functioning and Health (ICF) model [34]. From a legal standpoint, evaluations of work disability become more difficult due to stricter eligibility criteria with respect to objectivity, diagnosis, and prognosis of the disability. Sound support from evidence-based guidelines would, therefore, be welcome. The European Union of Medicine in Assurance and Social Security (EUMASS), a network of insurance medicine associations in seventeen European countries, recently published a comparison of work disability evaluation practices and the instruments in use, including guidelines [8]. This comparison was produced by several questionnaire rounds among central medical staff of participating countries. Two central questions in that study were

1 What is evaluated in your countries' work disability evaluation?
2 What instruments are used for these evaluations?

We were interested to determine what guidelines exist in different countries and their quality by focusing on the following research questions:

1. What *disease-oriented* guidelines have been developed to support the medical evaluation of work disability?
2. What is the quality of these guidelines in social insurance medicine?

Methods

1. Identification of disease-oriented guidelines to evaluate work disability.
We used the EUMASS table to determine the countries in which guidelines were reported to be in use. The Netherlands, the Czech Republic, Germany, the United Kingdom, and Switzerland were visited based on their reported use of the guidelines; no other countries had reported using guidelines for medical evaluations. The status of guidelines was assessed during the visits by determining if they were officially prescribed. Copies of the guidelines with explanation were collected. For this article we focused on the guidelines for evaluating work disability by SIPs that were prescribed by law or as an instruction by the ISI. We distinguished between disease-oriented guidelines (describing aspects of evaluations for certain pathologies) and process-oriented guidelines (describing aspects of evaluations, regardless

of pathology), a distinction that is evident from the relative guideline's title. We selected disease-oriented guidelines. To compare guidelines, we selected those that addressed the same diseases.

2. Quality appraisal of guidelines
The selected guidelines were scored using the AGREE instrument, which uses 4-point scales for each item: scope and purpose (3 items), stakeholder involvement (4 items), rigour of development (7 items), clarity and presentation of the recommendations (4 items), applicability of the guideline (3 items), and editorial independence (2 items). To correct for the different number of items in each domain, The AGREE instruments suggests calculating domain scores by relating the obtained scores (OS) to the maximum possible score (MaPS) and the minimum possible score (MiPS) using the formula

OS-MiPS/ MaPS-MiPS

As a test, one (Dutch) guideline (burnout) was scored by two researchers (WdB and DB) using the AGREE instrument and its user guide to establish if additional rules for scoring would be required. The test showed the need for additional scoring rules. We specified the clinical question and the target population and we adapted user guide item 11 (health benefits, side effects and risks) and 16 (options for management of the condition). The scoring rules we developed are presented in Appendix 1. The selected guidelines were then scored independently by two researchers (WdB and DB). The initial agreement between the researchers was determined using Kappa. Any differences were discussed, but if a difference remained, a decisive third researcher (JRA) would score as well, using the scores and arguments of the first two. We analysed the initial correlation between the two scoring researchers. As this use of the AGREE instrument is new in social insurance medicine, we asked one expert in each country who had participated in developing several guidelines for a reaction to our results: "Are these correct in your view and what is your explanation for any insufficient scores?"

Ethics committee:
This study was not submitted for ethical approval. The study included physicians who were not asked to perform specific professional actions for this study, but only to complete a questionnaire. All studied documents are in the public domain.

Results

1 Identification of disease-oriented guidelines to evaluate work disability. In Germany seven guidelines for SIPs turned out to be officially in use. In the Netherlands twenty-four were found and one in Switzerland. These guidelines are partly process-oriented and partly disease-oriented. Process guidelines were used in Germany (1), the Netherlands (8), and Switzerland (1). The German and Swiss guidelines each contain many recommendations that in the Netherlands are distributed over eight smaller guidelines. The recommendations refer, for example, to the relevance of the diagnosis for the evaluation and to the boundaries of the concept of disease. Another topic of these guidelines is the claimant's obligation to attempt to recover and find gainful employment. Yet another aspect is the relevance of distinguishing between the opinions of the claimant and the SIP. These recommendations represent the consensus of legal and medical experts on the principles of evaluation, but not on scientific evidence. These process-oriented guidelines were excluded.

Disease-oriented guidelines were in use in Germany (6) and the Netherlands (16), shown in Table 1. In the Czech Republic, a Barema-type of guideline is in official use, but this was excluded from this study as it evaluates impairments, not work disability.

Table 1: Diagnosis-oriented guideline for SIPs to country, publisher, and year of publication/revision, nr of pages (exc summary and addenda) and nr of references

Guideline (country and publisher)	Year	Pages	References
Aspecific Lumbar Disorder (NL, Health Council)	2005/ 2008	20	15
Myocardial Infarction (NL, Health Council)	2005/ 2008	22	45
Anxiety Disorders (NL, Health Council)	2007	30	27
Stroke (NL, Health Council)	2007	30	30
Breast Cancer (NL, Health Council)	2007	24	35
Chronic Fatigue Syndrome (NL, Health Council)	2007	26	22
Herniating Intervertebral Disc (NL, Health Council)	2007	20	14
Burnout (NL, Health Council)	2007	28	29
Depressive Disorder (NL, Health Council)	2007	32	29
Whiplash Associated Disorders (NL, Health Council)	2008	26	24
Arthritis Hip and Knee (NL, NVVG)	2007	28	56
Rheumatoid Arthritis (NL, NVVG)	2007	34	57
Chronic Obstructive Lung Disease (NL, NVVG)	2007	46	54
Chronic Heart Failure (NL, NVVG)	2007	30	41
Schizophrenia and associated psychoses (NVVG)	2007	50	135
Chronic Shoulder Disorders (NL, NVVG)	2007	21	37
Mental disorders (DE, DRV)	2001/ 2006	53	59

Herniating Intervertebral Disc (DE, DRV)	2002/ 2003/ 2005	26	28
Chronic Inflammatory Bowel Disease (DE, DRV)	2005	26	60
Coronary Heart Disease (DE, DRV)	2001/ 2005	20	54
Chronic Obstructive Lung Disease (DE, DRV)	2003/ 2005	34	47
Breast Cancer (DE, DRV)	2006	22	42

The Dutch guidelines, all implemented by law, were first developed by the Health Council of the Netherlands and later by the scientific association of SIPs (NVVG). The German guidelines were developed and prescribed by the German Institution of Social Insurance (DRV). The German guidelines were developed earlier than the Dutch and most have been updated since their inception.

2a. The appraisal of quality with the AGREE instrument of selected guidelines

Of the guidelines, four diseases were common to both countries: breast cancer, chronic obstructive lung disease, lumbar intervertebral disc herniation, and myocardial infarction.

The initial agreement between researchers was high for the Dutch guidelines (Kappa range 0.814-0.939), but low for the German counterparts (Kappa range 0.449-0.624). After discussing the different opinions of the researchers, agreement was reached on all items and scoring by the third researcher was unnecessary. The results are presented in Table 2.

Table 2. AGREE scores of selected guidelines to domain

	Breast Cancer		**Chronic Obstructive Lung Disease**		**Lumbar Intervertebral Disc Herniation**		**Myocardial Infarction**		**Total**	
	Dutch	German	Dutch	German	Dutch	German	Dutch	German	Dutch	German
Scope and purpose of the guideline	100	100	100	100	100	100	100	100	100	100
Stake-holder involve-ment	58	33	50	33	50	33	50	33	52	33
Rigour of development	10	19	19	24	19	19	14	29	16	23
Clarity and presentation	75	67	75	67	50	75	50	75	63	71
Applicability	11	0	11	33	0	0	0	0	6	8
Editorial indepen-dence	50	0	50	0	50	0	50	0	50	0

The scope and purpose of the guideline were well described in all eight guidelines; the score in both countries was **100%**. All guidelines were designed to support the medical evaluation of work disability by indicating what functional incapacities were to be expected in cases with a specific diagnosis.

Stakeholder involvement was **52%** for the Dutch and 33% for the German guidelines. Potential users were well defined (social insurance physicians), but the involvement of professional groups was found to be incomplete in seven of the eight guidelines. The patients' views were not sought in the German guidelines and only at the final stage in the Dutch. No guidelines were piloted among end-users before their publication.
Rigour of development was **16%** with the Dutch and **23%** with the German guidelines. How evidence was gathered and the scientific grounding of recommendations were not explicit in any guideline.
.
Clarity and presentation of the guidelines was **63%** for the Dutch guidelines and **71%** for the German. Although the recommendations were unambiguous and easily identifiable in almost all cases, they were not overly specific. Different options for assessing the condition of the guidelines were often mentioned, and the German guidelines provided tools for the evaluations.
Applicability scored **6%** in the Netherlands and 8% in Germany. Practical barriers and costs were not addressed in any guideline. The German guidelines contained indications of when to update them.
Editorial independence was limited in both countries. The Dutch guidelines reached **50%** on average as they were developed independently of the funding body, but with only a general procedure about conflicting interests. The German guidelines **(0%)** were developed entirely within the ISI and conflicting interests were not addressed.

2b Feedback on the AGREE scores by experts involved in developing several guidelines
The Dutch expert was involved in developing 11 of 16 then-published guidelines in the Netherlands and 3 of the 4 protocols that we scored on the AGREE instrument. He agreed to all our scoring after we discussed our scoring rules with him. He attributed low scores to the newness of creating guidelines for social insurance medicine in the Netherlands and that the short time allotted to create them was a factor. Stakeholder involvement was also reduced because patients' involvement was controversial in the beginning as there was concern about patients being biased with regard to the recommendations. The low figure on rigour of development was because the methods of development had not been recorded and because the field had no scientific tradition. The lack of specificity of the recommendations was due mainly to a lack of existing scientific research. Applicability scored low in the Netherlands as the guidelines were developed by the Health

Council, for whom this was not a regular activity. The aspects of applicability were considered by the ISI after publication of the guidelines.
The German expert was involved in developing five of six guidelines published at the time in Germany and in all the guidelines we scored on the AGREE instrument. He agreed to nineteen of the twenty-three scores after we discussed our scoring rules with him. Differences were due partly to how the German guidelines were described (experts involved were not identified with their specialisation) and to differences in the interpretation of items 13, 14, and 15. He commented that the development of guidelines was new in Germany and started from a need of the SIPs within the ISI, which explained the limited involvement of stakeholders. The involvement of patients' representatives was considered unhelpful because of expected bias. Testing among users was done implicitly as the guidelines were developed at the institution where the SIPs work. The selection of evidence and formulation of recommendations were carried out according to what the German experts considered the most important. No need had existed to document any more than they did for internal use, which accounted for the low score on the rigour of the guidelines' development. This internal development also accounted for the low score on applicability; this was included implicitly within the development process of internal guidelines. Editorial independence was not considered important, as the interests of the SIPs and the ISI were not supposed to conflict.

Discussion
In this study we looked for the existence of evidence-based guidelines for the medical evaluation of long-term work disability and the quality of development of these guidelines.

Main findings
Using the EUMASS comparison, we found guidelines for the medical evaluation of work disability, both disease- and process-oriented, in official use in four of seventeen European countries. In two of these countries we found twenty-two disease-oriented guidelines in official use in these evaluations. The AGREE instrument was applicable for scoring the selected Dutch and German guidelines, although minor adaptations to the AGREE instrument were necessary. Scoring German guidelines gave a smaller initial agreement than the Dutch, due to language problems and understanding of the German social insurance; however, the consensus procedure compensated for these issues. The guidelines scored well on 'scope and purpose' and 'clarity and presentation', and moderately on 'stakeholder involvement' in the Netherlands, but low in Germany; all guidelines scored low on 'rigour of development'. 'Editorial independence' and 'applicability' were low as a result of how production was organised.

Strengths and weaknesses
To our knowledge, this is the first study to identify and qualify medical

guidelines in social insurance medicine at an international level. As we were looking for official guidelines, we do not believe that we missed any in the countries we included; however, focusing on official guidelines may have resulted in finding fewer guidelines than are in practical use. For example, in Germany and Switzerland, guidelines are published by specialists in scientific journals. These are not in official use, but they may support physicians in their evaluations.

We used the AGREE instrument to determine the quality of the guidelines, which recommends using four appraisers for a good reliability [20]. Using a pilot procedure and two researchers for scoring, we obtained good agreement, which was supported by the opinion of the two experts who were involved in developing the guidelines. All items of the AGREE instrument proved to be relevant for testing the guidelines. We did not encounter important aspects that were not addressed by the AGREE instrument; further validation is needed however. Our adaptations are partly specifications of the scope of the AGREE instrument to the context of social insurance medicine, but are unlikely to influence the integrity of the AGREE instrument. Our adaptations of items 11 and 16 are less clear-cut translations that need to be tested.

Other studies
Our study corresponds with other research; the distinction between legal and medical guidelines fits with the results of Boer et al. [35] about the medical and legal aspects of a doctor's reasoning. The reliability of the AGREE instrument outside the clinical domain [16, 27, 28] was partly confirmed in our study, after minor alterations were made. Finding that guidelines do not fully meet the AGREE criteria is not uncommon [22, 36, 37, 38], partly due to the lack of a precise account of the development process and partly because of a lack of scientific evidence; both are not uncommon problems in drafting guidelines [22, 39, 40]. The relative lack of scientific research on the work participation of people with chronic diseases is also well documented [40, 41, 42, 43].

Impact
We found disease-oriented guidelines in only two participating countries, and there they are recent. Work disability is being evaluated on similar aspects in many countries, despite large differences in organisation of social insurance [6]; thus, we expect the development of guidelines to be likely elsewhere. Our results may be helpful in facilitating this.

Our comparison of development quality is based on four Dutch and four German guidelines, on four different pathologies. The German and Dutch social insurance systems differ in many aspects, but both require a medical statement about functional capacity in cases of claims for work disability benefit. From this perspective the guidelines are comparable in and

between countries. As the guidelines in these countries have been created in a similar fashion, we expect our results to be relevant to future disease-oriented guideline development in these countries.

We used the AGREE instrument as a tool for evaluating the quality of guideline development in social insurance medicine, a procedure that, to our knowledge, is new. It is unclear if using the AGREE instrument in a different domain is without problems; however, neither we, nor the experts we consulted, noticed any clear incongruence. The AGREE instrument is now being utilised in both Germany and the Netherlands.

With the AGREE instrument the quality of the development of guidelines can be scored, which is not the same as the quality of the recommendations. It is possible that the guidelines contain adequate recommendations that have been developed in a suboptimal way or whose development has been accounted for in a suboptimal way. Good practice, however, is best supported by guidelines that have been developed in a proven, optimal way. Several aspects need further consideration. The involvement of patients' representatives is now accepted in the Netherlands, after much discussion about the nature of their input; in Germany, however, this is not the case. This difference illustrates the ambiguity of the claimant's position in social insurance medicine: he is both passive object of the evaluation and participating subject in work disability. AGREE criteria are clear, however: participation of patients' representatives is mandatory. The development of the guidelines in the Netherlands has now been placed under the authority of the scientific association of SIPs, as this is viewed as the best way to retain independence from both the funding and implementing bodies. In Germany, financing, developing, and implementing within the ISI is considered effective, which illustrates the ambiguity of the profession of social insurance medicine as a discipline that needs to stress its independence and quality and a group of doctors working for administrative organisms with more interests than medical quality [29, 33]. AGREE criteria are clear on this aspect, too: a good guideline needs to be developed independently.

The inclusion of disease-oriented research into the practice of disability evaluation will help coordinate clinical, occupational, and social insurance medicine, in using the same concepts and findings, although in different spheres. The lack of scientific evidence may be compensated for, in part, by research on the aspects that influence disability with chronic conditions in general [41, 43]. Parallel to this, research needs to be commenced to establish if the guidelines actually contribute to quality improvement. Finally, the production of these guidelines will help formulate the questions that need to be addressed in future research to ground social insurance evaluations.

We expect that the diffusion of our results may aid further development of guidelines in social insurance medicine and, notably, help these become increasingly more evidence-based, which would assist in establishing a new and important mechanism for quality control in social insurance medicine. Paraphrasing Lohr [15], evidence-based evaluation practice in social insurance medicine would mean using the best evidence available and the best procedure possible to decide on the option that suits that claimant best.

Conclusion

Evidence-based guidelines form an important instrument for enhancing the quality of medical practice. Guidelines can provide a framework on which a clinician can ground diagnosis, therapy, and prognosis. Guidelines in social insurance medicine for the evaluation of work disability are a recent phenomenon, so far restricted to Germany and the Netherlands. We expect that disease-oriented guidelines can be useful in other countries as well, and can help the SIP ground his evaluation of capacity for work. For the practice of evaluating work disability, this would mean an important instrument to control quality. The AGREE instrument is suitably applicable for assessing the quality of guideline development in social insurance; nevertheless, some of the scoring rules need to be adapted to the context of social insurance. Existing guidelines do not meet AGREE criteria sufficiently. Notably, how patients' representatives can be involved and the editorial independence of the guideline developers need further discussion. The guidelines would profit from more specific recommendations and, for this, more research is needed on the functional capacity of people with disabilities. To date, research has focused primarily on the recovery from complaints, while mainly ignoring the resumption of work. The latter depends on much more than a health condition, but still, the challenge of health care should not only be to give relief for pain and suffering, but also to allow participation in society and to legitimise a disability benefit if needed for medical reasons.

Authors' contribution: WB designed the study, carried out the field work, and prepared the manuscript. DB participated in the scoring and drafting of the article. AR participated in the field work. PD supervised the field work and participated in drafting the article. HA supervised and participated in the drafting of the article.

Acknowledgements: The study was supported with a grant from the SIG Foundation, which had no involvement in the study itself or in the decision to submit the paper for publication. The authors wish to express their gratitude to Dr. Klaus Timner, MD, and Mr. Frans Westerbos, MD, who commented on the authors' scoring and interpretation of these scores.

Competing interests: All author(s) declare that they have no competing interests.

References

1 Organisation for Economic Co-operation and Development (OECD). **Sickness, disability and work: keeping on track in the economic downturn**. Background paper. Paris OECD 2009

2 Organisation for Economic Co-operation and Development (OECD) **Sickness, Disability and Work (Vol. 1): Norway, Poland and Switzerland**. Paris OECD 2006

3 Organisation for Economic Co-operation and Development (OECD) **Sickness, Disability and Work (Vol. 2): Australia, Luxembourg, Spain and the United Kingdom**. Paris OECD 2007

4 Organisation for Economic Co-operation and Development (OECD) **Sickness, Disability and Work (Vol. 3): Denmark, Finland, Ireland and the Netherlands**. Paris OECD 2008

5 Henderson, M. **Transformation of the personal capability assessment**. London Department of Work and Pensions 2006

6 Boer, WEL de; Besseling, JJM; Willems, JHBM. **Organisation of disability evaluation in 15 countries**. *Pratiques et organisation des soins* 2007 **38**: 205-217

7 Council of Europe. **Assessing disability in Europe**. Strasbourg Council of Europe,2002

8 www.eumass.com/media/taxonomy.pdf visited 5-9-2009

9 Wind H. **Assessment of physical work ability: the utility of functional capacity evaluation for insurance physicians**. PhD thesis Amsterdam Amsterdam University 2007

10 Überschär, 2008. **Quality assurance in the socio-medical assessment in the German Pension Insurance**. *[in German] Gesundheitswesen* 2008 **70**: 690-695

11 Verbeek, JHAM Van Dijk, FJH. **Assessing the ability to work**. *BMJ* 2008 **336**: 519-520

12 Burgers, JS; Cluzeau, FA; Hanna, SE; Hunt, C Grol, R. **Characteristics of high-quality guidelines: evaluation of 86 clinical guidelines developed in ten European countries and Canada**. *Int J Technol Assess Health Care* 2003 **19**: 148-57

13 www.nice.org.uk; www.aezq.de; www.cbo.nl. visited 5-9-2009

14 Sackett, DL; Rosenberg, WMC; Gray, JAM; Haynes, RB; Richardson, WS. **Evidence based medicine: what it is and what it isn't**. *BMJ* 1996 **312**: 71-72

15. Lohr, KN. **The quality of practice guidelines and the quality of health care**. In Selbmann, HK: *Guidelines in health care: report of a WHO conference*. Baden Baden Nomos Verlaggesellschaft 1998, 42-52

16 Cates, JR; Young, DN; Bowerman, DS; Porter, RC. **An independent AGREE evaluation of the Occupational Medicine Practice Guidelines**. *Spine* 2006 **6**: 72-77

17 Schaafsma, F; Hugenholtz, N; Boer, A de; Smits, P; Hulshof, C; Dijk, F van. **Enhancing evidence-based advice of occupational physicians**. *Scand J Work Environ Health*, 2007 **33**: 368-378

18. Weide, WE; van der, Jos; Verbeek, HAM; van Dijk, FJH. **Relation between indicators for quality of occupational rehabilitation of employees with low back pain**. *Occup Environ Med* 1999 **56**: 488–493

19 Nieuwenhuizen, K; Verbeek, JH; Siemerink, JC; Tummers-Nijsen, D. **Quality of rehabilitation among workers with adjustment disorders according to practice guidelines; a retrospective cohort study**. *Occ Env Med*. 2003 **60**: suppl 1:i21-i25

20 AGREE collaboration. **Appraisal of Guidelines for Research & Evaluation**. www.agreecollaboration.org, 2001 visited 5-9-2009
21 AGREE Collaboration: **Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project**. *Qual Saf Health Care* 2003 **12**: 18-23
22 Muth C. Gensichen; J. Beyer; M. Hutchinsons; A. Gerlach FM. **The systematic guideline review: method, rationale and test on chronic heart failure**. *BMC Health Services Research* 2009 **9**: 74
23 Burgers, JS; Baily, JV' Klazinga, NS; Van der Bij, AK; Grol, R; Feder, G. **Inside guidelines: Comparative analysis of recommendations and evidence in diabetes guidelines from 13 countries**. *Diabetes Care*, 2002 **25**: 1933-1939
24 Burgers, JS. **Guideline quality and guideline content: are they related?** *Clin Chem*, 2006 **52**: 3-4
25 Appleyard, TL; Mann, CH; Khan, KS. **Guidelines for the management of pelvic pain associated with endometriosis: a systematic appraisal of their quality**. *BJOG* 2006 **113**: 749-57
26 Nagy, E; Watine, J; Bunting, PS; Onody, R; Oosterhuis, WP; Rogic, D; Sandberg, S;, Boda, K; Horvath, AR; IFCC Task Force on the Global Campaign for Diabetes Mellitus. **Do guidelines for the diagnosis and monitoring of diabetes mellitus fulfil the criteria of evidence-based guideline development?** *Clin Chem*, 2008 **54**: 1872-82
27 Hulshof, C Hoenen J. **Evidence based practice guidelines in OHS: are they agreeable?** *Ind Health* 2007 **45**: 26-31
28 Manchikanti, L; Singh, V; Helm, S. II; Trescot, AM; Hirsch, JA. **A critical appraisal of 2007 American College of Occupational and Environmental Medicine (ACOEM) Practice Guidelines for Interventional Pain Management: an independent review utilizing AGREE, AMA, IOM, and other criteria.** *Pain Physician* 2008 **11**: 291-310
29 Waddell, G Aylward, M. **The scientific and conceptual basis of incapacity benefits**. Norwich The Stationary Officer 2005
30 Aylward, M. **Origins, practice and limitations of disability assessment medicine**. In: Halligan PW, Bass C, Oakly DA. *Malingering and illness deception*. London OUP 2003 pp 287-299
31 de Boer, WEL; Wind, H; van Dijk, FJH; Willems, JHBM. **Interviews for the assessment of long-term incapacity for work: a study on adherence to protocols and principles**. *BMC Public Health* 2009 **9**: 169
32 Lipsky, M. **Street level bureaucracy**. New York Russell Sage 1980
33 Berendsen, L. **Bureaucratic dramas**. PhD Thesis *[In Dutch]* Tilburg University Tilburg, 2007
34 World Health Organisation (WHO). **International Classification of Functioning, Disability and Health (ICF)** Geneva WHO 2001
35 de Boer, WEL; Brage, S; Donceel, P; Rus, M.; Willems, JHBM. **Medicolegal reasoning**. *BMC public Health* 2008 **8**: 335
36 Tulder MW van, Tuut M, Pennick V, Bombardier C, Assendelft WJ. **Quality of primary care guidelines for acute low back pain**. *Spine* 2004 **29**: E357-62
37 Boluyt, N; Lincke, CR; Offringa M. **Quality of evidence-based pediatric guidelines**. *Pediatrics* 2005 **115**: 1378-91

38 Poitras, S; Avouac, J; Rossignol, M; Avouac, B; Cedraschi, C; Nordin, M; Rousseaux, C; Rozenberg, S; Savarieau, B; Thoumie, P; Valat, JP; Vignon, E Hilliquin, P. **A critical appraisal of guidelines for the management of knee osteoarthritis using Appraisal of Guidelines Research and Evaluation criteria**. *Arthritis Res Ther* 2007 **9**: R126

39 Ketola, E; Kaila, M; Honkanen, M. **Guidelines in context of evidence**. *Qual Saf Health Care* 2007 **16**: 308-12

40 Slebus, FG; P. Paul F. M. Kuijer, J. (Han) H. B. M. Willems, Monique H. W. Frings-Dresen and Judith K. Sluiter. **Work ability in sick-listed patients with major depressive disorder**. *Occ Med* 2008 **58**: 475-479

41 Dekkers-Sánchez PM, J L Hoving, J K Sluiter, M H W Frings-Dresen. **Factors associated with long-term sick leave in sick-listed employees: a systematic review** *Occup Environ Med* 2008 **65**: 153-157

42 Oxman, AD; Schünemann, HJ; Fretheim, A. **Improving the use of research evidence in guideline development**: **16. Evaluation**. *Health Res Policy Syst*, 2006 **8**: 28

43 Slebus, FG; PPFM Kuijer; J (Han) H B M Willems, J K Sluiter, M H W Frings-Dresen, **Prognostic factors for work ability in sicklisted employees with chronic diseases** *Occup Environ Med* 2007 **64**: 814–819

Appendix 1:

AGREE from clinic to social insurance:
Most AGREE items were found to be directly applicable to the guidelines in social insurance medicine. We specified the AGREE items 2 and 3 and adapted items 11 and 16 from a clinical context to a context of evaluation in social insurance medicine.

Scope and purpose

1 The overall objective(s) of the guideline is (are) specifically decribed.
2 The clinical question(s) covered by the guideline is (are) specifically described.

We specified the clinical question as: what functional incapacities are to be expected with diagnosis X?

3 The patients to whom the guideline is meant to apply are specifically described.

We specified the target population as: the claimants to whom the guidelines should be applied.

Stakeholder involvement

4 The guideline development group includes individuals from all the relevant professional groups.
5 The patients' view and preferences have been sought.
6 The target users of the guideline are clearly defined.
7 The guideline has been piloted among target users

Rigour of development

8 Systematic methods were used to search for evidence
9 The criteria for selecting the evidence are clearly described.
10 The methods used for formulating the recommendation are clearly described.
11 The health benefits, side effects and risks have been considered in formulating the recommendations.

We considered the risks of following or not following of the guideline as criterion.

12 There is an explicit link between the recommendations and the supporting evidence.
13 The guideline has been externally reviewed by experts prior to its publication.
14 A procedure for updating the guideline is provided.

Clarity and presentation

15 The recommendations are specific and unambiguous
16 The different options for management of the condition are clearly presented

We decided to score if the guideline indicates differences in recommendations in different situations.

17 Key recommendations are easily identifiable
18 The guideline is supported with tools for application.

Applicability

19 The potential barriers in applying the recommendations have been discussed.
20 The potential cost implications of applying the recommendations have been considered
21 The guideline presents key review criteria for monitoring and/ or audit purposes.

Editorial independence

22 The guideline is editorially independent from the funding body.
23 Conflicts of interest of guideline development members have been recorded

Chapter 6:
Interview protocols in social insurance medicine

W.E.L. de Boer[1], J.H.L.Wijers[2], J. Spanjer[3], I. van der Beijl[1], W. Zuidam[1], A. Venema[1]

[1] Department of Quality of Life TNO Hoofddorp the Netherlands
[2] WOSM company. Zeist the Netherlands
[3] UWV Groningen the Netherlands

Abstract

Background
The quality of the disability evaluation interview – the standard tool for assessing claims in the Netherlands under the Invalidity Insurance Act (WAO) – leaves much to be desired. Pressure to produce a validated and reliable evaluation will be all the more urgent if the plans of the present government are implemented. In the past, attempts have been made to design a valid instrument in the form of interview protocols.

Methods
This article identifies the existing protocols and examines their underlying principles; it also considers their similarities and differences. The three protocols considered are the Disability Assessment Structured Interview (DASI), the Interview of Methodical Assessment (IMA) and the Multi-Causal Analysis (MCA). The similarities and differences of the three protocols were examined in terms of their underlying principles and practical technique.

Results
All were found to be based on the experience of the interviewers and none were appropriately validated. All three are based on a complex of legal, medical and psychological considerations. The similarities between the protocols were found to be greater than the differences. In this article, the principles used to design the protocols are discussed in relation to the research literature. They are underpinned primarily with indirect arguments. Further development and practice-based research could significantly raise the degree of professionalism in this field.

Introduction

The evaluations performed by insurance physicians are often the subject of criticism [1]. Yet as far as we are aware, the nature of the problem has never been researched: it may be due to faulty procedures, or to evaluations being unreasonably severe or – conversely – much too lenient. Several studies have plausibly argued that evaluations lack consensus [2,3,4,5]. In general, an evaluation is based on pre-information, an interview, a physical examination and sometimes further tests or examinations. In everyday practice, the medical aspect of WAO disability evaluation depends heavily on the interview with the claimant [6,7,8]. It is striking that there is so little scientific underpinning for this aspect of the evaluation.

It seems probable that in the future, the evaluation interview will come under greater pressure in relation to claims under the Work and Income according to Labour Capacity Act (WIA) [9]. Although reintegration reports (including information from specialists) will play a more major role in evaluation than is currently the case, the question about the completeness and permanence of the disability will still be met chiefly by information obtained in the interview [10]. So there is an urgent need for a robust examination instrument in the form of a validated interview protocol of proven effectiveness. However, no such instrument is available to date.

Attempts have been made to design interview protocols for WAO purposes [11,12,13,14,15]. By 'interview protocols' we mean schematic rules for conducting interviews, which in essence prescribe the topics to be covered and also the interviewing technique to be employed. The interview protocols currently used in WAO settings are not based on scientific theory. Equally, little research has been carried out into their effectiveness, validity or reliability. Spanjer [17,18] investigated the reproducibility of WAO evaluations, and their inter-rater and intra-rater reliability. He found that evaluations based on a video-presentation of the Disability Assessment Structured Interview were highly reliable in a laboratory setting, but probably less so in practice. In his opinion, physicians did not probe sufficiently with their questions, and paid too much attention to medical matters and disease. The interview protocols are an integral part of the evaluations. Their status is indicated in the recommendations on insurance medicine by the Health Research Council of the Netherlands (RGO) [8]: they are information-gathering models linked to verification models. Two instances of the latter are described in the RGO's recommendations: the argument-based claim evaluation and the insurance medicine reference framework. We found no evidence of interview protocols based on the insurance medicine reference framework. Nor did we find evidence, either in the literature (PubMed) or via our own research sources, of argument-based protocols being used outside the Netherlands for disability claim evaluation [19].

In this article we consider the following research questions:
What are the similarities and differences between the published protocols?
What are the protocols based on?

Method
By studying descriptions of the published protocols and interviewing their designers, we extrapolated the characteristics, basic principles and claims of the various protocols and set them out in a way which enabled comparison (research question 1.) These interviews were conducted in several rounds. First, we and all the designers jointly drew up a set of concepts, which were subsequently elaborated in bilateral interviews. Afterwards, the results were jointly discussed and checked. This method of working was necessary because the designers each had their own terminology. In practice, there was occasionally a greater degree of consensus than reflected in the written descriptions.

Having found similarities and differences between the various underlying principles, we attempted to identify the basis of each (research question 2). For each basic principle, we examined the literature (Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde, PubMed and Psychinfo), looking for theories and protocols which might serve to develop and test the three interview protocols.

Results of research question 1: What protocols are in use?
The first interview protocol to be introduced into insurance medicine and into the training curriculum for insurance physicians was the Interview of Methodical Assessment (IMA) [11,13]. A second type is the Disability Assessment Structured Interview (DASI) [14], which was developed in response to the IMA: the latter was seen as inflexible and insufficiently focused on the Functional Capacity List (FCL). The third type is the Multi-Causal Analysis (MCA) [12,15], which was also designed in response to the IMA's perceived lack of flexibility and in order to deal with the perceived difficulty of claimants to verbalize their claim effectively at the beginning of the interview. To sum up:
In insurance medicine, there are three published protocols for disability evaluation interviews.
All three protocols are based on legal, medical and psychological considerations.
Parts of the protocols are different, but there are also considerable similarities.

Table 1 presents a systematic comparison of the main features of the protocols. For further details, the reader is referred to the original publications cited.

Table 1 Comparison of interview protocols in insurance medicine

	Disability Assessment Structured Interview (DASI)	**Interview of Methodical Assessment (IMA)**	**Multi-Causal Analysis (MCA)**
Typical characteristics	*Probe for factual, detailed examples of restrictions/ limitations & capacities* Tight structure All topics must be addressed but sequence is free	*Frequent use of summaries* Address all 9 topics. Same sequence at start of each item (work, claim and claim-related ailments, alternative work) Ask claimant to give account of 'normal day'** Tight structure *Ignore claimant's spontaneous remarks* *** *Correct claimant's inadequate answers* ****	*Empathize with claimant; claimant must feel heard (good communication essential!)* *All 6 topics must be addressed but sequence is free and switching is allowed between topics* *Probe as much as possible and pick up on any remarks by claimant*
Topics	*Work* *Disease data* *Limitations & restrictions experienced* *Activities/ handicaps* *Claimant's opinion (suited to current work or lighter work?)* *Physical examination* *Physician's opinion*	*Claim items* *- work* *- claim & claim-related ailments* *- alternative work* Verification items* *-motivation* - convictions about cause of disease and handicap - fitness & vitality *- change, mental/personal* - life events - future - physical examination	- Health - *Work* - *Private life* - Functioning - Person - Physical examination (Claimant's opinion emerges in all these topics)
Introduction	Explain purpose of interview Explain procedure Summarize case Remove any resistance	Explain purpose of interview *Agree on procedure (first give opinion on present situation and then go on to other issues)* Discuss interview agenda	Briefly explain procedure Briefly summarize case Quickly engage claimant in conversation
Most important feature	All topics must be addressed	*All topics must be addressed, using at least 1 leading question*	All topics must be addressed
Ending	Ask if claimant has anything to add *Physician's conclusion and possible consequences* Explain further procedures	Ask if claimant has anything to add Final summary Conclusion Follow-up appointments, if any	Refer back to claimant's restrictions & limitations Offer claimant chance to add anything; pick up on new remarks or unfinished business Agreements and future planning

Structure	Tight structure, especially in section on 'limitations and restrictions'	Tight structure for first three claim items	Loose structure
Duration	30-45 minutes	45 minutes	30-45 minutes
Verification (consistency & plausibility of argument-based claim evaluation)	*Elicit factual, detailed examples (NB consistency) Consistency between all interview topics; also between claimant's self-report and other information* Common sense Insurance physician must be convinced	*Achieved through protocol's structure and probing questions by insurance physician*	*Gather information by asking and probing on all topics Consistency in claimant's self-report Supplementary information may be obtained from third parties*
Use of topic summaries as ...	interviewing technique	*fixed part, concluding each topic (essential to protocol)*	interviewing technique
Sequence of topics	Preferably as prescribed in model, but still at physician's discretion	Preferably as prescribed in model, but still at physician's discretion	*Free, providing all topics are addressed. Switching allowed between topics.*

Notes:
Highly typical characteristics are indicated in bold italics.
Interpretations and interview styles may vary from one physician to another.

* Verification items: these items may be used by the insurance physician to verify the reliability and consistency of the claimant's story.
** Normal day: description of a normal day (e.g. the day before the day of the evaluation interview) in the life of the claimant.
*** Spontaneous remarks: claimant's remarks that are not directly related to the topic currently under discussion.
**** Inadequate answers: claimant's responses that do not directly relate to the question asked but deviate from it.

Interview of Methodical Assessment (IMA)

The IMA's main underlying principle is the argument-based claim evaluation [8]: the premise is that the claimant's claim and arguments must be verified and underpinned, with the emphasis on his activity limitations and capacities. The responsibility for underpinning the claim, for incapacity for work and for behaviour towards recuperation rests primarily with the claimant. For this reason, every effort is made in the interview to put the claimant on an equal footing with the insurance physician, and the claimant is tested to see if he is able and willing to bear the responsibility. The interview is semi-structured, with comprehensiveness as its key characteristic. The conversation proceeds according to rules, with a set sequence of questions and a set role for the claimant. Any deviations, such as spontaneous remarks by the claimant, are considered as a result of the test. There are 12 topics to be addressed.

The description of the method contains many detailed instructions on how to ask specific questions and how to interpret answers. Every topic is concluded with a summary. Together, the summaries serve as building-bricks for the overall conclusion. The verification criteria are plausibility, internal consistency, and congruity with insights from medicine and the social sciences. The insurance physician delivers an overall conclusion, underpinning arguments, and an evaluation. The objective is to arrive at a systematic evaluation based as far as possible on verifiable facts.

Disability Assessment Structured Interview (DASI)
Here too, the main underlying principle is argument-based claim evaluation [8]. There is a major focus on the differences between the claimant's current state and pre-morbid state as operationalizations of disease or impairment. Another key role is played by factual and detailed examples that the claimant gives (preferably without prompting) of every activity he performs and of the restriction of capacity that he claims to experience. These examples serve to reduce possible malingering or aggravation by the claimant and to identify residual capacity for work. The interview is semi-structured: the topics are fixed by the insurance physician but they may be addressed in any sequence. Once all the information has been gathered, the insurance physician forms a judgment, which he then clearly states to the claimant. The purpose of this method is to reach a systematic evaluation, focusing on what actually needs to be assessed in practice.

Multi-Causal Analysis (MCA)
This interview protocol, too, follows the principle of argument-based claim evaluation [8]. However, in the MCA, the emphasis is on mapping impairment and motivation factors in the claimant. The protocol therefore includes questions about psychological and social aspects, as well as medical. A relationship of trust is vital between the insurance physician and the claimant. The insurance physician should display an attitude of empathy, respect and interest; he should probe where necessary and take due account of the claimant's subjective perceptions.

The MCA is a heuristic approach which gives considerable freedom to the insurance physician. The topics are fixed but are set out in broad categories. The interview is loosely structured: it may be dynamic, switching from one topic to another on spontaneous cues from the claimant. Probing questions reveal the plausibility and consistency of the claimant's report and the degree of his incapacity. Any aspect of the claim may be discussed at any point in the interview. The main focus is on the claimant's perception, so it is essential for the interviewer to keep probing. The purpose of this method is to reach an understanding evaluation.

In practice, the three protocols may not be strictly applied. We know from the literature that assessing physicians tend to be pragmatic in using the

available instruments [3,20]. However, that is a subject beyond the scope of this article.

Similarities between the protocols

All three protocols build on the principle of argument-based claim evaluation.
All three protocols seek to identify functional capacities as well as activity limitations.
The truth criteria are: plausibility (is the information submitted in the claim likely to be based in fact?) and consistency (not contradicted in any way by the claimant, and congruent with insights from medicine and the social sciences).
In all three protocols, the evaluation is medical, but it also takes non-medical aspects into account and is therefore multifactorial.
All three protocols are semi-structured, to a greater or lesser degree.

Differences between the protocols
The interview has a different character in each protocol: the IMA is like a test, the MCA approximates to a dialogue and the DASI resembles a questionnaire.
Spontaneous remarks and inadequate answers are handled differently: in the IMA and MCA the physician asks more probing questions, whereas in the DASI, he steers the conversation back to the semi-structure by using summaries.
Empowerment is another point of difference, i.e. the physician builds a relationship of trust with the claimant so that the interview becomes a joint activity, enabling the claimant to play an active part in the evaluation and come up with solutions to his health problems. The IMA and MCA seek to empower the claimant, whereas the DASI does not.

The IMA seeks to build up a thorough and detailed picture before a judgment is formed. In the DASI, the assessor's aim is to identify impairments and capacities, which takes up relatively less time than in the IMA. The MCA is somewhere between the other two protocols on this point.
The aims differ: in the DASI, the aim is to fill in the Functional Capacity List (FCL); other protocols also aim to help the claimant to gain self-understanding, promote his functional restoration and encourage him to get back to work.

The similarities between the protocols appear to be based on certain underlying principles, which form the basis for the second research question.

Results of research question 2: What are the protocols based on?
The similarities and differences between the three protocols derive from their designers' views about legislation and science. These views are not made explicit in the development and description of the protocols, nor are they underpinned by scientific argument. The descriptions are strongly based on practical experience. All the interview protocols share common underlying principles and the differences between them are mainly a matter of degree. However, none of the protocols has ever been actually validated. On the basis of legislation and scientific knowledge, the following points may be made about the acceptability of the principles themselves:

The designers all start from the principle of the argument-based claim evaluation, i.e. assessing a claim on the basis of arguments submitted by a claimant. This fits in with the statement in the RGO's recommendation that this is the only approach to have generated instruments for evaluation of work disability [8].
All three protocols link in with the modern notion that an evaluation should be based primarily on positive capacities, besides negative impairments. This is also mandatory for the working method of the Employee Insurance Schemes Implementing Body (UWV), which uses the Functional Capacity List.
The truth criteria (plausibility and consistency) are in line with the Evaluation Decree [21], and before it, the Guidelines on Medical Incapacity for Work (MAOC) [22].
All three protocols represent a multifactorial approach, as opposed to a purely monocausal (medical) one. This explains the variety of topics addressed in the various protocols. All the topics are relevant as pointers towards the claimant's functioning [21,23], but the extent of the topic categories and the various options they contain have not yet been subjected to testing. For instance, what is covered by the topic 'Motivation' in the IMA? To what extent does it overlap with 'Impairments experienced' in the DASI or 'Person' in the MCA? There is a substantial overlap in all the protocols as regards the topics to be addressed. Of the three, the IMA is the most detailed and the MCA is least specific.
All three protocols devote attention to the position of the claimant during the interview. The claimant must substantiate his claim (explain why he is no longer working, or only partly working) and collaborate on the evaluation. The assessor must verify whether the information submitted by the claimant in support of the claim is correct, and whether it should be amended or supplemented. It is generally accepted that these positions have an impact on the obtaining of information and the forming of a judgment. For instance, the claimant's report may understate or exaggerate his ailments, impairments or capacities – in good or bad faith [24]. The exact nature of the impact on the information-gathering process has not yet been established through empirical research: Misleading behaviour due to illness may be common, or it may be rare [25]. Moreover, if somebody

says he cannot work, what is the reason behind this? Is he afraid that working will damage his health? Does he think it unreasonable that he is expected to work in his current circumstances? Or is it simply that he does not want to work? The three protocols all refer to these positional features, for instance in the scripting of the assessor-assessee relationship in the introduction, the list of topics to be addressed and the manner in which they are to be addressed.

According to the protocols, the relationship between the claimant and the insurance physician is to be managed by specific techniques. This depends, among other things, on how each one views his task and role in the interview. Is it based on a classical doctor-patient relationship (especially in the MCA) or is the claimant supposed to be a talkative individual with aims of his own (especially in the MCA), who resolves a problem with a doctor [26]? The latter is preferable when the aims of the assessor and the claimant are the same. This is least stressed with the DASI: the greater the difference of opinion between the two about the outcome of the evaluation, the more likely that the evaluation will turn into a conflict of interest that will have to be handled through conflict behaviour.

The key role part played by conflict resolution is also reflected in the emphasis given to this topic in the training, education and protocols that pertain to this professional field. In the literature, the relational component is offered as a possible explanation for the differences between assessors [7,8,9]. For instance, poor interaction with the claimant may incline the assessor to think that the claimant's symptoms or impairments are difficult to objectify. The three protocols display no significant differences on this point.

All the interviews can be interpreted as ways of testing what the claimant sees as his capacities. From a legal point of view, this is a sound argument, since anybody who makes a claim must be able to substantiate it. Another argument is that the claimant is the person best acquainted with his own situation. Two recent doctoral dissertations [5,28] have shown that the reliability of self-report is debatable. The value of self-report in incapacity evaluations has not yet been established by empirical research, so for the insurance physician, it is a problematic starting-point. Another argument in favour of verifying the claimant's self-report is that the claimant will be more inclined to put his capacities to good use if he feels his voice has been heard during the evaluation interview and that he has made a positive contribution to the outcome. This principle is confirmed in the literature on empowerment: the greater the claimant's participation in decision-making, the more his empowerment increases [29]. It has not yet been established whether this also applies to disability claim evaluation.

All three protocols feature a combination of semi-structuring and, where appropriate, probing. Both are found to a varying degree. Semi-structuring is an obvious option for a number of reasons. Firstly, it is an efficient way of addressing the core of a problem, assuming the assessor is generally familiar with the nature of the problem. Secondly, it prevents the assessor

from falling into his own trap: for instance, his extensive past experience may cause him to jump to a conclusion so that he misses unexpected information or overlooks contradictory details, and then proceeds to form a judgement [30]. Thirdly, semi-structuring is one way of making cases comparable: by always using the same set of questions, one is soon alerted to any discrepancies in the claimant's self-report. This is the reason why the IMA is a kind of test.

The protocols display differences in interpretation of the tasks. It is always debatable how extensive an evaluation should be in order to be described as 'sufficient'. All three protocols state the aim of being both efficient and comprehensive, each in its own way. However, the descriptions reveal that the IMA is the most detailed, while the DASI is the most strategic. How this works in practice, and on what basis the one might be preferred to the other, is unknown.

Another difference in interpretation of the tasks has to do with whether the assessor should solely focus on evaluating the claim or also aim to help the claimant further. The DASI focuses on the former, whereas the MCA emphatically targets the latter.

Discussion

In comparing the interview protocols that play such a key role in disability evaluation, we find similarities and differences. The fact that the protocols can now be compared and are therefore susceptible to evaluation, may be seen as a step forward. It is striking that such an evaluation has never taken place in the past. It is also noteworthy that the theoretical and empirical basis of the protocols is so implicit. This prompts the view that evaluations are conducted in a 'black box' and that the claimant may not have sufficient opportunity to exercise his rights. For the sake of the professional status of insurance medicine, it is desirable that the evaluation interview instrument should be implemented in a transparent and fully validated manner. At present, this is not sufficiently the case. That does not mean that doctors play their interviews by ear: there is no evidence of this from the protocols. For instance, the medical approach and interpretation of the law seem to be sound and defensible in all the protocols. Nevertheless, there is considerable divergence between notions from the behavioural sciences and those from insurance medicine. In practice, the relational aspects of the protocols and the degree of structuring may make a significant difference. Some of the principles, such as the value of certain interviewing techniques in the opinion-forming process, lend themselves to empirical study. The debate about the interpretation of tasks touches on dilemmas that are inherent to the profession, and are best managed through consensus [31,32]. This means that social insurance physicians must continually adapt to the socio-political reality of their day and age [33].

Conclusions

1 There are similarities and differences between the three protocols.
2 None of the protocols is the best of the three; in fact, no single protocol has proven to be sound.
3 The underlying principles - plausibility and consistency as truth criteria, relational fastidiousness, the multifactorial approach and semi-structuring – all of these are acceptable, but they require empirical underpinning for disability evaluation settings under the terms of the Invalidity Insurance Act.
4 Several of the differences between the protocols (such as whether the claimant should not only be assessed but also activated, and whether the evaluation is efficient) raise questions about the valid concerns of insurance medicine, and what these ought to be. However, this is a matter of requiring a consensus within the profession as a whole and in consultation with the UWV.

Recommendations
We think that the interviews based on these protocols can help to make insurance medicine more professional. The three protocols are based on relatively acceptable basic principles. However, it would be desirable to test out those principles in the insurance medicine context so that professional practice is scientifically underpinned. In parallel to this, insurance medicine practitioners could develop general criteria for the evaluation interview, which would not only be a matter of good science but also of creating consensus on the depth, efficiency and standards which are desirable for the profession. Further research into the effects of the differences between the protocols, as well as into their practical use by insurance physicians, is also desirable. It is doubtful whether a single model will ever be rated as ideal, but in the near future, it should certainly be possible to give disability evaluations a sounder and more scientific basis than they have at present.

References

1 Committee on Labour capacity. **Turning laopur capacity into work**. *[in Dutch]* Doetichem Elsevier 2001

2 Razenberg PPA. **Judgment makin in social insurance medicine**. *[in Dutch]* Amsterdam PhD Thesis UvA 1999

3 Bont AA de, Berendsen L, Boonk MPA, Brink JC van den. **At the doctor's office**. *[in Dutch]* Zoetermeer CTSV 2000

4 Kerstholt JH, Boer WEL de, Jansen EM. **Disability assessments: effects of response mode and experience**. *Disab & Rehab* 2006 **28**: 111-115

5 Brouwer S. **Disability in chronic low back pain**. Groningen PhD Thesis RU Groningen 2004

6 Landelijk Instituut Sociale Verzekeringen (Lisv). **Guideline on methods of evaluation**. *[in Dutch]* Amsterdam Lisv 2000

7 Mul CAM, Verklei HEM, Hazelzet AM, Roos F de. **Quality of formation of judgment**. *[in Dutch]* Hoofddorp TNO 1999

8 Raad voor Gezondheidsonderzoek (RGO). **Advice on insurance medicine**. *[in Dutch]* Den Haag RGO 2004

9 Willems JHBM, Kroneman H. **Physician or solicitor: new act of work disability insurance changes the role of the social insurance physician**. *[in Dutch] Medisch Contact* **58**: 337-340 2003

10 Boer WEL de, Brenninkmeijer V, Besseling JJM. D**isability evaluation abroad: lessens for the new act on work disability insurance**. *[in Dutch]* SMA **59**: 408-413 2004

11 Boer WEL de, Duin JA, Herngreen H. **Manual for Interview of Methodical Assessment**. *[in Dutch]* Utrecht NSOH 1997

12 Boer WEL de, Meijers JM, Verhoeff M. **The multicausal analysis: theory and practice**. *[in Dutch]* Hoofddorp TNO 2000

13 Herngreen H, Boer WEL de. **Manual for Interview of Methodical Assessment**. *[in Dutch]* Utrecht SSG 1988

14 Spanjer J. **Disability assessment structured interview**. *[in Dutch]* Amsterdam Lisv 2000

15 Wijers JHL. **Handicap and integration**. *[in Dutch] Handicap en beleid* 1996 ***3***: 20-22

16 Molen HT van der, Kluytmans F, Kramer F. **Conversation: skills and models**. *[in Dutch]* Groningen Wolters Noordhoff 1995

17 Spanjer J. **The reproducibility of evaluation of legal work disability**. *[in Dutch] TBV* **8**: 195-198 2000

18 Spanjer J. **The inter and intra rater reliability of evaluations of legal work disability**. *[in Dutch] TBV* **8**: 234-241 2000

19 Boer WEL de, Brenninkmeijer V, Zuidam W. **Long term Disability arrangements**. Hoofddorp TNO 2004

20 Horstman K, Berg M, Plass S, Heusden M van. **Guidelines and the Legal character of social Insurance**. *[in Dutch]* Amsterdam Lisv 1997

21 **Assessment Decree of july 8th 2000** *[in Dutch] Staatscourant* 307 2000

22 Landelijk Instituut Sociale Verzekeringen. **Guideline medical criterion of work disability**. *[in Dutch]* Amsterdam Lisv 1996

23 World Health Organisation (WHO). **International Classification of Functioning and Health**. WHO Geneva 2001
24 Aylward M. **Origins, practice and limitations of disability assessment medicine**. In: Halligan PW, Bass C, Oakly DA. *Malingering and illness deception*. Londen OUP pp 287-299 2003
25 Halligan PW, Bass C, Oakly DA. **Malingering and illness deception**. Londen OUP 2003
26 Meershoek A. **Back to work again**. *[in Dutch]* Amsterdam Thela Thesis 1999
27 Nauta, A. **An eye for an eye**. *[in Dutch]* Groningen PhD Thesis RU Groningen 1996
28 Reneman MF. **Functional capacity evaluation in patients with chronic low back pain: reliability and validity**. Groningen PhD Thesis RU Groningen 2004
29 Itzhaky H, Schwartz C. **Empowering the disabled: a multidimensional approach**. *Int J Rehabil Res* 1998 **21**: 301-10
30 Pennington N, Hastie R. ***The story model for juror decision making***. In: Hastie R, editor. *Inside the Juror; The psychology of juror decision making*. New York Cambridge University Press 1993 p 192-221
31 Teulings CN, van der Veen RJ, Trommel W. **Dilemma's of social security**. *[in Dutch]* Den Haag VUGA 1997
32 Weel ANH; Kelder MJ; Nauta AP. **Practice dilemmas for occupational physicians and social Insurance physicians**. *[in Dutch]* Houten BSL 2005
33 Stone D. **The Disabled State**. London Mac Millan 1985

Chapter 7:
Interviews for the assessment of long-term incapacity for work: a study on adherence to protocols and principles

Wout EL de Boer [1], Haije Wind [2], Frank JH van Dijk [2], Han HBM Willems [2]

[1] Department of Quality of Life, TNO, Hoofddorp, the Netherlands
[2] Academic Medical Centre, University of Amsterdam, Coronel Institute of Occupational Health, Amsterdam, the Netherlands

Abstract

Background
Assessments for long-term incapacity for work are performed by Social Insurance Physicians (SIPs) who rely on interviews with claimants as an important part of the process. These interviews are susceptible to bias. In the Netherlands three protocols have been developed to conduct these interviews. These protocols are expert- and practice-based. We studied to what extent these protocols are adhered to by practitioners.

Methods
We compared the protocols with one another and with the ICF and the biopsychosocial approach. The protocols describe semi-structured interviews with comparable but not identical topics. All protocols prescribe that the client's opinion on his capacity for work, and his arguments, need to be determined and assessed. We developed a questionnaire to elicit the adherence SIPs have to the protocols, their underlying principles and topics. We conducted a survey among one hundred fifty-five experienced SIPs in the Netherlands.

Results
Ninety-eight SIPs responded (64 %). All respondents used some form of protocol, either one of the published protocols or their own mix. We found no significant relation between training and the use of a particular protocol. Ninety percent use a semi-structured interview. Ninety-five percent recognise having to verify what the claimant says and eighty-three percent feel the need to establish a good relation ($p=0.019$). Twelve topics are basically always addressed by over eighty percent of the respondents. The claimant's opinion of being fit for his own work or other work, and his claim of incapacity and his health arguments for that claim, reach a hundred percent. Description of claimants' previous work reaches ninety-nine percent.

Conclusions
Our study shows professional consensus among experienced Dutch SIPs about the principle of assessment on arguments, the principle of conducting a semi-structured interview and the most crucial interview topics. This consensus can be used to further develop a protocol for interviewing in the assessment of incapacity for work in social insurance. Such a protocol can improve the quality of the assessments in terms of transparency and reproducibility, as well as by enabling clients to better prepare themselves for the assessments.

Background:

People at work get sick every now and then, generally for a short time. A minority of these remains sick for a longer time and some are forced to turn to social insurance. Arrangements for people with long-term incapacity for work exist in social insurance in many countries, among which the Netherlands and the UK. In these schemes, a benefit is possible for those insured that meet the legal criterion of being permanently unable to gain sufficient income because of illness or handicap [1]. This meets the requirements of what Gordon [2] called the 'handicapped role', or 'disability role' according to Waddell and Aylward [3]. That concept describes the health condition of the person as 'disabled', his rights to be (partly) exempt from work, his obligation to look for cure and rehabilitation, and his obligation to look for work that may still be fit for him. The legal criteria are formulated in abstract terms, which facilitate tailor-made assessments of people in very different circumstances [4]. In order to be granted a benefit, insured people have to file a claim and they have to be assessed. These assessments lead to conclusions about the residual capacity for work of the claimant in terms of the scheme of disability benefit. Between countries there is considerable variation in social insurance schemes, but the assessments of long-term incapacity for work are most often performed by specialised social insurance physicians (SIPs) [5]. This is, for example, the case in the Netherlands and the UK. The quality of these assessments is unknown. One might consider the degree of work resumption to be an indicator of the quality of the assessments as they predict the claimant's capacity for work. However, work resumption alone is not a valid quality indicator as it is influenced by the personal factors of the claimant (e.g. motivation, attitudes and beliefs, social factors) and by factors on the labour market. Relating incapacity for work only to objective medical findings would do an injustice to claimants as (in-) capacity for work is a relational concept that requires the consideration of work factors as well [3]. In several countries process indicators and expert based guidelines have been developed to support the work of the SIPs [6,7]. In one type of guideline, the profession of the SIPs makes clear how they consider that assessments should be done according to diagnostic categories [6,8]. In another type of guideline, prescriptions are provided about how to perform the assessments in general [6,9,10].

SIPs may use a number of sources to acquire information for their assessments. The first source is the claimant, who has knowledge of his situation and needs to have the opportunity to explain his claim and his arguments, and so to put forward grounds on which his claim is to be evaluated. Interviews are, therefore, crucial and they can be organised either face to face with the SIP (as is the case in the Netherlands and in the UK) or through an intermediate professional such as a medical specialist (e.g. Germany: [11]) or a social insurance officer (e.g. Sweden: [12]). Apart from the claimant, the SIP may also request information from the treating

physician, the employer and external medical experts. Social insurance physicians in the Netherlands mainly base their judgement about the work ability of claimants on the information they receive from the claimants [13]. One might argue that the claimant's opinion of what he can and cannot do in work should be sufficient evidence on which to provide a benefit [14,15,16]. The claimant's opinion, however, may be governed by coping problems and economical interests and so the claimant may be biased [17,18]. Furthermore, the legal criteria for benefit for long-term incapacity for work are abstract [4] and it is unsure if claimants have a good understanding of the assessment criteria. So the interviews with the claimants are not only meant to be used for listening, but also to inform the claimant and to verify the claim against the legal criteria [19].

It is unknown how these interviews are conducted in practice. Guidelines for assessment of incapacity for work indicate what needs to be addressed in the interview. They do not indicate how this needs to be done – whether it is in a free conversation, following a form or using some structure. It seems, therefore, likely that every SIP develops his own routine, guided by his education, his experience and his preferences. This is not without risks: several studies show substantial differences in results between assessors, which underlines that these interviews do not meet criteria of reliability [20,21,22]. Both for society and for claimants, it is hard to accept that the final outcome of an assessment is not only depending of the physical or mental condition, but also on the person who performs the assessment.

A protocol that describes how to conduct a reliable interview to assess incapacity for work would be of value for both SIPs and claimants. Structured interviews are known to enhance the reliability of information gathering and conclusion [23]. In the Netherlands, three interview protocols have been drafted from practice to be used in the assessments. Based on these protocols, the profession has the opportunity to develop standards of good social insurance medicine. These protocols and their underlying principles provide an opportunity to study the professional consensus about these interviews. SIPs in the Netherlands receive, depending on where they work and get their education, training in one or several of these protocols, and they are free to use them or to adapt them to the SIPs' own wishes. This situation provided an opportunity to find out if there is professional consensus in practice on how to conduct the interviews. For this reason, we were looking for an answer to the following question:

To what extent are SIPs familiar with the protocols and to what extent do they adhere to the principles of the protocols?

Methods:

Design:
The design of the study is a descriptive survey among social insurance physicians.

Participants and recruitment procedures:
A total of one hundred fifty-five social insurance physicians (SIPs) were sent a questionnaire. These SIPs were selected from the nine hundred members of the Dutch Association of Insurance Medicine (NVVG). These one hundred fifty-five SIPs had earlier pronounced their commitment to contribute to the development of social insurance medicine. They had volunteered to participate in pro deo projects of their association to professionalise their work. All were working in disability evaluation for the Dutch Act on Insurance of Incapacity for Work (WAO).

Protocols:
In the Netherlands, three protocols to perform disability assessment interviews have been published, all based on practical experience: the Interview of Methodical Assessment (IMA: [24]), the Disability Assessment Structured Interview (DASI: [25]) and the Multi Causal Analysis [26]. Boer et. al. [27] report on a comparison of the protocols. For a detailed description of the protocols, see Appendix 1.
The protocols all describe semi-structured interviews, indicating the topics that need to be addressed during the interviews and their sequence. To a varying extent, the protocols describe the techniques or procedures of the interview such as the introduction, summaries and ending. All protocols are based on the principle of assessment on arguments [28], which means that the opinion of the claimant of his capacity and incapacity and his arguments for that opinion are to be discussed, completed if necessary and verified. This verification first takes place in the interview itself by comparing the claimant's opinion with other information such as facts regarding the past and future and his experiences other than in work. Furthermore, the SIP considers medical records, physical examinations, the history of sick leave, and return to work activities in order to form his opinion on the claimant's capacities. Finally, all protocols pay attention to the special context of social insurance, which makes the interviews different from medical examinations in health care [3, 9, 10, 28]. The protocols prescribe a critical attitude for the SIPs and suggest special attention for the introduction to the interview in which a clarification of the purpose and procedure is explained to the claimant. The protocols do not describe conditions for interviewing such as time, the qualifications of SIPs or an optimal moment of assessment. The topics that address a claimant's disability can be compared to ICF [29] and a biopsychosocial approach [3], and can be said to match both. See Table 1 for this comparison.

Table 1 Interview protocols according to ICF and biopsychosocial approach (BPS)

ICF	BPS	IMA	DASI	MCA
Disease	Bio.	Health complaints, cause of disability	Information on disease.	Health and disease.
Impairments.	Bio.	Health complaints that prevent claimant from working. General health.	Information on disease.	Health and disease.
Activity limitations.	Bio.	Health complaints that prevent claimant from working. Activities of daily living.	Actual functioning.	Actual functioning.
Participation problems.	Social.	Claimant's perception of his capacity for own work. Claimant's perception of his capacity for other work.	Claimant's perception of his capacity to do his own or other work. Actual problems of participation.	Actual functioning.
Personal factors.	Psycho.	Motivations.	Perceived burden in the work.	Person.
Environmental factors.	Social.	Work description.	Work description.	Work description. Private situation.

The protocols show differences as well. The IMA provides the most detailed description of twelve topics in a fairly strict sequence. The DASI is less strict and uses the Listing of Functional Capacities (LFC) as a checklist, together with six other topics in a preferred sequence. The LFC is the output form in use at the Dutch Institute of Social Insurance and indicates six clusters of activities that are relevant for functioning in work. In the DASI, the claimant is asked to give examples of his actual functioning. The MCA is the least strict, providing five areas of conversation that need to be explored in a preferred sequence.

The topics of the different protocols resemble each other but are not precisely the same. The topics are partly medical such as 'Medical history' or 'General health', but also psychosocial such as 'Private situation', 'Motivation' or 'Life events'. Topics cover the experiences and events of the past, examples of which are 'Medical history' and 'Life events', and the present such as 'Claimant's opinion of his actual capacity for work' and expectations for the future.

The IMA invites the claimant to follow precisely the questions asked and not to elaborate on personal associations. Summaries in IMA are not only used as an interview technique but also as formal stepping stones for the conclusion. The DASI invites the claimant to describe his functioning with actual examples from everyday life. Summaries are used as an interview technique. The MCA strives to achieve maximal trust from the claimant by quickly focussing on the aspects that bother the claimant. Thus, it is expected that the claimant will open up and present his capacities and incapacities in an honest manner. Summaries are used as an interview technique – they are utilised to encourage the claimant's participation by showing that the SIP understands what the claimant says.

Procedure and set-up:
The authors formulated a list of questions on four subjects to investigate the research question cited above. The description of the protocols was used to draft the questionnaire. The questionnaire was mailed to the participating SIPs.

I. The first subject was the familiarity and the use of the protocols by the SIPs. The respondents were asked if they used one or more of the three protocols and if they had been trained in these. The answer could also be that they did the assessment their own way, not using any of the protocols.
II. The second subject was the direction of the interview in the situation of social insurance. The respondents were asked who decided on the topics of the interview and their sequence. The answers could be that the interview was structured, that there was an application of a sequence of topics, that the SIP or claimant determined the topics, and whether specific examples of limitations of activities were asked. The answers were categorised over the three protocols, a combination of these protocols, or labelled as 'own protocol'.
III. The third subject was the position of the claimant towards the SIP. Respondents were asked (1) if they always checked the information provided by the claimant, and (2) if having a good relationship with the patient during the assessment is important. The answers were categorised over the three protocols, a combination of these protocols, or labelled as 'own protocol'.
IV. The fourth question was about the topics that the SIPs basically always address during the disability assessment. A list of topics was proposed, based on the protocols. The answers were categorised over the three protocols, a combination of these protocols, or labelled as 'own protocol'.

Data analyses:
The number of participating SIPs, mean age, and years of experience were noted. The application of a protocol and having been trained in it were noted in percentages of the SIPs. The answers to the second and third questions were noted in percentages of SIPs, in total and per protocol. The answers to the fourth question were noted in the frequency of topics that are basically always addressed, in total and by protocol. Differences between the groups of SIPs concerning questions 1, 2 and 4 were tested using T-tests. A p-value < 0.05 was considered to be statistically significant. The answers to question 3 were analysed using the exact two-sided McNemar test, considering a p-value <0.05 to be statistically significant. As the primary aim of the survey was to describe the relation between SIPs' familiarity with the protocols and their adherence to those protocols, and not to determine the causes of adherence, no multivariate analyses were performed.

Ethics committee:
This study was not submitted for ethical approval. The study includes physicians who are not asked to perform specific professional actions for this study but to fill in an anonymous questionnaire.

Results:

Of the hundred and fifty-five SIPs, ninety-eight returned a completed questionnaire (64 %). Sixty-four SIPs (64 % of 98 respondents) were male and the average age was 47.7 years (SD=6.9). Sixty-six had more than 10 years' experience in disability evaluation based on the Dutch Act on Insurance of Incapacity for Work (WAO). We have no information on non-respondents.
Respondents were asked if they were trained in one or more of the three protocols and if they used them. Eighty-seven percent of the respondents were trained in IMA, forty percent in DASI and twenty-seven percent in MCA. All respondents used some form of protocol: twenty-three percent reported to use IMA, twelve percent DASI and twenty-two percent MCA, whilst forty-two percent reported to have constructed their own mix. We found no significant relationship between the training received and the use of a particular protocol.
Respondents were asked who determined the topics of the interview – the claimant or the SIP – and, if applicable, in what sequence. The results are shown in Table 2.

Table 2 Direction of the interview in total and by use of protocol, % yes

	● Total	'use IMA'	'use DASI'	'use MCA'	'use several'	'use own model'
N:	99	23	12	22	20	22
%:	(100%)	(23%)	(12%)	(22%)	(20%)	(22%)
Interview follows a fixed pattern (N=97)	90%	95%	100%	77%▼	95%	86%
Use fixed sequence of items (N=99)	63%	70%	75%	50%	70%	55%
Items are determined by claimant (N=99)	0%	0%	0%	0%	0%	0%
Items first by claimant then by SIP (N=99)	39%	22%▼	42%	50%	45%	41%
Ask specific examples of limitations of activities (N=99)	75%	65%	75%	86%	80%	68%

Percentages are column percentages and are tested with the Pearson Chi-square test. The contrast is: 'subgroup' vs. 'other cases'. ▲ and ▼: $p<0.05$ for significantly high and low percentages. Symbols are based on significance only, not on Effect Size. Tests and symbols refer to horizontal comparisons.

Ninety percent of the SIPs have their interview structured and sixty-three percent of the SIPs structure by applying a fixed sequence of topics as prescribed by the protocols. The others maintain structure on a more abstract level than on topics, indicating fields of discussion such as 'Private situation' or 'Person'. With none of the respondents the topics were determined by the claimant, but for thirty-nine percent of the SIPs, the claimant may have some room for his own topics at the start of the interview, after which the SIP takes over. Asking for specific examples of limitations of activities is done by seventy-five percent of the SIPs. The use of interview protocols affects only two aspects: a fixed pattern is less reported by users of MCA and leaving room for the claimant to start with his own topics is less seen with IMA. This is in agreement with these protocols.
SIPs were asked about their professional attitude towards the interviews. There are significantly more SIPs who recognise their role in having to verify what the claimant says (95%), than there are SIPs who recognise the need of establishing a good relationship (83%, $p<0.02$, McNemar's test). Between users of a particular protocol, there are no significant differences.
SIPs were asked if the introduction to the interviews has a specific function. The results are shown in Table 3.

Table 3 Attitude towards the interview in total and by use of protocol, % yes

	• Total	'use IMA '	'use DASI'	'use mca'	'use several '	'use own model'
N:	97	22	12	22	19	22
%:	(100%)	(23%)	(12%)	(23%)	(20%)	(23%)
Need to put the client at ease (N=96)	77%	57%▼	50%▼	91%	89%	86%
Need to clarify the interview purpose (N=97)	94%	100%	100%	86%	84%	100%
Need to clarify the interview procedure (N=96)	61%	81%▲	50%	50%	68%	55%

Percentages are column percentages and are tested with the Pearson Chi-square test. The contrast is: 'subgroup' vs. 'other cases'. ▲and ▼: $p<0,05$ for significantly high and low percentages. Symbols are based on significance only, not on Effect Size. Tests and symbols refer to horizontal comparisons.

Clarifying the purpose of the interview is common amongst ninety-four percent of the respondents. The need to put the client at ease is recognised by seventy-seven percent and significantly less so by users of IMA and DASI. Users of MCA, of a combination of protocols and of their own protocol try to break the ice significantly more than those using IMA and DASI exclusively. Users of IMA were most keen on instructing the claimant about the procedure of the assessment with eighty-one percent, which is significantly higher than the sixty-one percent of the whole group.
Respondents were asked to name the topics they basically always address.

These are shown in Table 4.

Table 4 Topics that are addressed in total and by use of protocol, % yes

		'use IMA'	'use DASI'	'use MCA'	'use several'	'use own model'
N:	• Total 99	23	12	22	20	22
%:	(100%)	(23%)	(12%)	(22%)	(20%)	(22%)
Claimant's opinion fit for own work or other (N=99)	100%	100%	100%	100%	100%	100%
Claim and health arguments (N=99)	99%	100%	100%	100%	95%▼	100%
Work (N=99)	99%	100%	100%	95%	100%	100%
Perceived limitations of activities and obstacles (N=99)	97%	91%	100%	100%	100%	95%
Actual functioning (N=99)	94%	96%	100%	91%	100%	86%
Medical history (N=99)	95%	96%	100%	91%	100%	91%
Private situation (N=99)	93%	91%	100%	95%	80%▼	100%
Activities/ handicaps (N=99)	91%	87%	83%	91%	100%	91%
Conclusion SIP (N=99)	90%	100%	100%	86%	90%	77%▼
Future (N=99)	88%	83%	92%	82%	95%	91%
General Health (N=99)	82%	87%	92%	77%	95%	64%▼
Possibility to do other work (N=99)	86%	87%	83%	82%	90%	86%
Motivation (N=99)	68%	70%	50%	77%	70%	64%
Life-events (N=99)	67%	65%	58%	68%	70%	68%
Change mentally, as a person (N=99)	62%	70%	58%	64%	70%	45%
Person (N=99)	59%	52%	33%	68%	80%▲	50%
General health (N=99)	49%	43%	50%	55%	70%▲	32%
Causes of disability (N=99)	44%	52%	17%▼	41%	65%▲	36%

Percentages are column percentages and are tested with the Pearson Chi-square test. The contrast is: 'subgroup' vs. 'other cases'. ▲ and ▼: $p<0,05$ for significantly high and low percentages. Symbols are based on significance only, not on Effect Size. Tests and symbols refer to horizontal comparisons.

Twelve topics are mentioned by over eighty percent of the respondents and six topics by between forty-four and eighty percent of the respondents. The claimant's opinion of his being fit for his own work or other work and his claim of incapacity and the health arguments he has for that claim stand

at hundred percent. A description of the claimant's previous work reaches ninety-nine percent. The claimant's opinion of the 'causes of his disability' and his 'general health' do not reach an agreement of fifty percent of the respondents.

Discussion:

In this study, we examined the extent to which three Dutch interview protocols for the assessment of incapacity for work and their underlying principles were known and adhered to. The respondents were a selected group of experienced SIPs who were all doing assessments for the Dutch Act on Insurance of Incapacity for Work (WAO) in the Netherlands and motivated for professional development (N=155).

Main findings
Ninety-eight SIPs responded to the questionnaire. They were all trained in at least one of the protocols. Fifty-eight percent reported to use one of these and forty-two percent had constructed their own protocol. We found no significant relation between being trained in a protocol and using it. This corresponds with the finding that a single element of training without control on implementation does not yield stable results [30, 31, 32]. The results also indicate that SIPs do make their own mix of recommendations that are given by the different protocols. Respondents agreed on the idea of conducting a semi-structured interview, most often by using a fixed sequence of predefined topics. The protocols define eighteen topics altogether, twelve of which are basically always addressed by over 80% of the respondents. The SIPs recognised their position of having to verify what the claimant says and to make an effort to get good cooperation with the claimant rather than establishing a good relationship. Semi-structured interviews can lead to a more reliable gathering of information by using a construct of what is being assessed [23]. All protocols, although using loosely defined topics, can be said to use an implicit concept of disability. This concept matches the ICF and the biopsychosocial approach, both being recognised as authoritative in this field. All protocols aim at determining not only limitations but also capacities, which is in accordance with modern opinions about the participation of people with disabilities [33].
The context of assessment in social insurance implies the need for a fair trial and a critical attitude of the SIP [19]. A fair trial requires among others that the claimant must be invited to state his claim and his arguments. It is a professional choice to assess on the basis of this claim and arguments rather than to determine disability only on presumed objective medical findings. It is unsure however, how valid and reliable a claimant's opinion of his situation is and how he reports this during claim assessment [214, 34, 35].

Strengths and weaknesses
This study reports the expert opinion of SIPs whose daily work it is to conduct interviews for the assessment of incapacity for work. The SIPs are not representative of all SIPs as they are selected on their ambition to contribute to their profession. With regard to adherence to the protocols they are probably a positive selection. The SIPs were all trained in one or more of the protocols and had had the opportunity to develop a protocol that served their daily needs. We asked the SIPs for their opinions on principles of interviewing in assessment of incapacity for work but we do not know how they perform in practice. It is uncertain to what extent protocolled interviews address ICF fields in an even manner in practice. Slebus et. al. [36] and Brage et..al. [37] found that in assessment of incapacity for work personal factors and environmental factors were less addressed than the other fields of ICF.

Impact
Our results open up a new way of quality control of the assessments by using a protocol to conduct the interviews. As the basic principles are accepted by the majority of SIPs their application can be assessed. It is possible to repeat this study in other countries to find the common principles that SIPs apply in different arrangements. That may make it possible to develop interview protocols elsewhere too. It seems likely that interview protocols need to be tailored to the arrangement at hand. Long term incapacity for work may need a different protocol from short term incapacity or for allowances for other handicaps. In any case further scientific testing is needed to establish more than face validity. The degree to which interviews in assessment of incapacity for work would best be structured is not known. Full structuring is not likely to be possible as many topics may be relevant in a specific case but there is no evidence to decide on what topics are the most relevant in all cases [38, 39, 40]. In order for such protocols to be effective they need to be implemented and applied in practice. Our study shows that earlier protocols were not blindly followed after training and we did not study why this was the case. It needs to be proved that a protocol that parts from accepted basic principles will do better. Some form of follow up after training will probably be necessary [31].

Conclusions
One way of dealing with the susceptibility to bias of assessment interviews is to use protocols for interviewing the claimants. In Dutch practice several such protocols have been developed. These protocols correspond with concepts in the ICF and the biopsychosocial approach. Our study indicates that there is professional consensus among experienced Dutch SIPs about the principle of assessment on arguments, the principle of conducting a semi-structured interview and the most crucial interview topics. Crucial topics cover all fields of ICF. This consensus can, without striving for a detailed and universally applicable protocol, be used to further develop

professional consensus on SIPs attitude, structuring of the interview and the selection of relevant topics that are more precisely circumscribed and based on evidence about what constitutes disability. This consensus can provide a starting point for further validation and development of a new protocol that can be implemented in practice and evaluated. It would need more than a single training in order to really be implemented. Some form of control is necessary.
If such a protocol is developed, implemented and controlled, the quality of the assessments would be improved in terms of transparency and reproducibility. The assessments would also become more comparable which would make them more accessible to scientific research on behaviour of both the SIP and the claimant. It would also enable claimants to better prepare themselves to the assessments which would make their position more equal to that of the SIP. The transparency of the reports and the satisfaction of the claimants would be endorsed by this.

Competing interests: All author(s) declare that they have no competing interests.

Authors' contribution: Wout EL de Boer designed the study, did the field work and prepared the manuscript. Haije Wind participated in drafting the manuscript. Frank JH van Dijk and Han HBM Willems supervised the project and participated in drafting and revising the manuscript.

Acknowledgements: The study was supported with a grant from the SIG Foundation who had no involvement in the study itself or in the decision to submit the paper for publication

References

1 Council of Europe. **Assessing disability in Europe**. Strasbourg CoE 2002

2 Gordon G. **Role theory and illness**. New Haven College and university press 1966

3 Waddell G, Aylward M. **The scientific and conceptual basis of incapacity benefits**. Norwich TSO 2005

4 Mabbet D, Bolderson H, Hvinden B. **Definitions of disability in Europe: a comparative analysis**. Uxbridge Brunel University 2002

5 Boer WEL de, Besseling JJM, Willems JHBM. **Organisation of disability evaluation in 15 countries**. *Pratiques et organisation des soins*. 2007 **38**: 205-217

6 Health council of the Netherlands. **General introduction to the protocols for insurance medicine**. *[in Dutch]* 2006/22 Den Haag Gezondheidsraad 2006

7 Legner R, Cibis W. **Quality control in socio-medical assessment**. *[in German] Rehabilitation* 2007 **46**: 57-61

8 Becker E, Horn S, Hussla B, Irle H, Knorr I, Korsukewitz C, Pottins I, Rohwetter M, Schuhknecht P Timner K. **Guidelines for the sociomedical assessment of performance in patients suffering from discopathy or associated diseases**. [in German] *Gesundheitswesen* 2003 **65**: 19-39

9 Bundesversicherungsanstalt für Angestellte, Landesversicherungsanstalten, Bundesknappschaft, Bahnversicherungsanstalt und Seekasse im Verband Deutscher Rentenversicherungsträger **Medical assessment for legal disability insurance**. *[in German]* 2001

10 Riemer Kafka G (ed) **Assessment in social insurance medicine**. *[in German]* Basel Schweizerische Ärzteverlag AG 2007

11 Reck R. **Chronic back pain and expert opinion**. *[in German] Versicherungsmedizin* 2005 **57**: 171-177

12 Ydreborg B Ekberg K, Nilsson K **Swedish social insurance officers' experiences of difficulties in assessing applications for disability pensions--an interview study** *BMC Public Health*.2007 **7**: 128

13 Bont AA de, Berendsen L, Boonk MPA, Brink, JC van den. **At the social insurance physician's office**.*[in Dutch]* Zoetermeer CTSV 2000

14 Chang YC, ChenSea MJ, Jang Y, Wang JD. **A simple self-rating assessment model of residual work capability for occupational permanent disabilities**. A*merican journal of industrial medicine* 2000 **38**: 539-547

15 Fleten N, Johnsen R, Forde OH. **Length of sick leave - why not ask the sick listed?** *BMC public health* 2004 **4**: 46

16 Norrby E, Lindahl I. **Reliability of the instrument DOA: dialogue about ability related to work**. *Work* 2006 **26**: 131-139

17 Mendelson G & Mendelson D. **Malingering in the medicolegal context**. *Clin J Pain* 2004 **20**: 423-432

18 Bianchini KJ, Greve KW, Glynn G. **On the diagnosis of malingered pain-related disability : lessons from cognitive malingering research**. *Spine* 2005 **5**: 404-417

19 Aylward M. **Origins, practice and limitations of disability assessment medicine**. In: Halligan PW, Bass C, Oakly DA. *Malingering and illness deception*. Londen: OUP 2003 pp. 287–299

20 Kerstholt J, Boer WEL de, Jansen NJ. **Disability assessments: effects of response mode and experience**. *Disabil Rehabil* 2006 **28**: 111-115
21 Brouwer S, Dijkstra PU, Stewart RE, Goeken LN, Groothoff JW, Geertzen JH. **Comparing self-report, clinical examination and functional testing in the assessment of work-related limitations in patients with chronic low back pain**. *Disabil Rehabil* 2005 **27**: 999-1005
22 Spanjer J, Krol B, Brouwer S, Groothoff JW. **Interrater reliability in disability assessment base don a semi-structured interview report**. *Disabil Rehabil* **30**: 1885-1890
23 Conway JM, Jako RA, Goodman DF. **A meta-analysis of interrater and internal consistency reliability of selection interviews**. *J Appl Psych* 1995 **80**: 565-579
24 Boer WEL de, Duin JA, Herngreen H. **Manual for interviewing in methodic assessment** *[in Dutch]* Utrecht SSG 1997
25 Spanjer J. **Disability assessment structured interview**. *[in Dutch]* Amsterdam Lisv, 2000.
26 Boer WEL de, Meijers JM, Verhoeff M. **Multicausal analysis: theory and practice** *[in Dutch]* Hoofddorp TNO 2000
27 Boer WEL de, Spanjer J, Wijers JHL. **Interview protocols in social Insurance medicine** *[in Dutch] TBV* 2006 **14**: 17-23
28 Health Council of the Netherlands. **Assessment, treatment and coaching** *[in Dutch]* 2005/10 Den Haag Gezondheidsraad 2005
29 World Health Organisation (WHO). **International Classification of Functioning, Disability and Health (ICF)**. Geneva WHO 2001
30 Kok R, Hoving JL, JHAM Verbeek, Schaafsma FG, Smits PBA, Vlek FJM, Duin JH, Dijk FJH van. **Evidence based Insurance medicine: evaluation of a workshop on EBM**. [article in Dutch] *TBV* 2008 **16**: 381-387
31 O'Brien MA, Rogers S, Jamtvedt G, Oxman AD, Odgaard-Jensen J, Kristoffersen DT, Forsetlund L, Bainbridge D, Freemantle N, Davis D, Haynes RB, Harvey E. **Educational outreach visits: effects on professional practice and health care outcomes.** *Cochrane Database of Systematic Reviews 2007*, Issue 4. Art. No.: CD000409. DOI: 10.1002/14651858.CD000409.pub2
32 Davis D. Does CME work? **An analysis of the effect of educational activities on physician performance or health care outcomes**. [Review] *Int J Psychiatry* 1998 **28**: 21-39
33 Organisation for Economic Co-operation and Development (OECD). **Transforming Disability into Ability. Policies to Promote Work and Income Security for Disabled People**. Paris OECD 2003
34 McCahon S, Strong J, Sharry R, Cramond T. **Self report and pain behaviour among patients with chronic pain**. *Clin J Pain* 2005 **21**: 223-231
35 Reneman MF, Jorritsma W Schellekens JM Goeken LN. **Concurrent validity of questionnaire and performance based disability measurement in patients with chronic nonspecific low back pain**. *J Occup Rehab* 2002 **12**: 119-129
36 Slebus FG Sluiter JK, Kuijer PP, Willems JH, Frings-Dresen MH.**Work-ability evaluation: a piece of cake or a hard nut to crack?** *Disab and Rehab* 2007 **29**: 1295-300
37 S Brage; P Donceel; F Falez. **Development of ICF core set for disability evaluation in social security** *Disabil Rehabil* 2008 **30**: 1392-1396

38 Franche R.L. Krause N. **Readiness for return to work following injury or illness**. *Journal of occupational rehabilitation* 2002 **12**: 233-256

39 Hunt DG, Zuberbier OA, Kozlowski AJ, Berkowitz J, Schultz IZ, Milner RA, Crook JM, Turk DC. **Are components of a comprehensive medical assessment predictive of work disability after an episode of occupational low back trouble?** *Spine* 2002 **27**: 2715-2719

40 Nusselder WJ, Looman CW, Mackenbach JP. **Nondisease factors affected trajectories of disability in a prospective study**. *J Clin Epidemiol* 2005 **58**: 484-494

Appendix 1:

Description of three disability interview protocols, procedure and topics

Interview of Methodical Assessment (IMA)

1 The IMA protocol describes ten topics that need to be addressed by at least one question and more if the SIP thinks the topic to be relevant to the case he is handling. The topics are clustered into topics that permit the claimant to state his claim and topics that permit the SIP to check on plausibility and consistency. The topics are best addressed in the sequence of the protocol and for the first three topics, this is mandatory. The description of the IMA contains many detailed instructions on how to ask specific questions and how to interpret answers. This enables the SIP to draft a complete picture of the claimant in his situation.

2 The IMA protocol requires a precise introduction, in which the aim and procedure of the assessment are explained and in which the SIP stresses that the claimant's opinion of his actual situation is of great importance and that the opinion of other people (for example, the treating physician) and events of the past will be dealt with later on during the interview. The claimant is asked to agree with these rules. Thus, the SIP introduces rules for the interview that challenge the claimant to show his self-consciousness and autonomy. This enables the SIP to see if the claimant is able to follow these rules.

3 A physical examination, if necessary, is scheduled after the interview.

4 After each topic, a summary is given by the SIP and after the entire interview, a general summary is given. After each summary, the claimant is invited to comment on it. At the end, the SIP gives his provisional opinion and explains the further procedure.

Claim items:
- Work description: Would you please describe the work you used to do?
- Claimant's perception of his capacity for own work: Do you think you could do that work now, fully or partly? If not, what do you experience in your health that prevents you from doing it?
- Claimant's perception of his capacity for other work: Do you think you could do other work? What would that need to look like?

Items to check:
- Motivation: How do/ did you like doing the work you used to do?
- Claimant's perception of the cause of disease and handicap: What do you think to be the cause of your being ill and disabled?

- General Health: Were you generally healthy and fit before you became disabled?
- Changes (mental, personal): Would you say you have changed as a person over the past period of sick leave?
- Life – events: Did you experience important events in the years before you reported sick? Which?
- Claimant's perception of the future: What do you expect about your future health/ work situation?
- Activities of daily living: Could you please describe an ordinary day, e.g. yesterday and indicate what you did, how you managed that and whom you met, in a chronological order?
- Physical Examination is scheduled at the end of the assessment.

3) Conclusion of the SIP, for the moment, is relative to the claimant's opinion.

Disability Assessment Structured Interview (DASI)

1 The SIP is focused on the differences between the pre-morbid state and the actual state that indicates disease. Another key role is played by concrete and detailed examples that the claimant gives or is asked to give of every activity he performs and of the restriction of capacity that he claims to experience. This serves to reduce possible malingering or aggravation by the claimant. These examples are used to identify residual capacity to work. A semi-structured interview is conducted in which topics are fixed by the SIP but their sequence is free. All topics must be discussed, preferably in order of the protocol, but the SIP can decide to do otherwise. The description of DASI does not give examples of questions but considerations as to why and how the different topics are of importance. The purpose of this method is to reach a systematic assessment of what is to be assessed – the claimant's capacity for work.

   DASI has a strong structure; in particular, in topics 3 and 4 the SIP asks for concrete and detailed examples, which must be consistent and plausible.
   Ask further information through others (treating physician, employer etc.).

2 In DASI the SIP explains the purpose of the assessment and the procedure. The SIP summarises the claimant's record. Putting the client at ease, the SIP explains the aim of the assessment.

3 A physical examination is scheduled after the interview.

4 At the end, the SIP states clearly his opinion of the claimant's capacities.

1. Work description and perceived burden in the work (motivation and consistency).
2. Medical history and information on disease: complaints, cause, treatment (impairments).
3. Claimant's perception of (in-) capacity in examples, if needed, with help of LFC (restrictions of activities).
4. Actual functioning and problems of participation: current activities and relationships (focus on capacities).
5. Claimant's perception of his capacity to do his own or other work (claimant's position in the assessment).
6. Physical examination (consistency and plausibility).
7. Opinion of the SIP.

Multi Causal Analysis (MCA)

1 MCA is designed to help the SIP to determine the causes of restricted functioning and so to be able to give suggestions to promote a return to work. The approach is biopsychosocial and the disability is primarily conceived of as behaviour. The instruction describes general principles, fields of discussion and the relevance of these. The emphasis is put to the claimant's motivation and hindrances he experiences. The psychological and social aspects are determined as well as medical aspects. All subjects must be discussed but the order is free.

2 The SIP briefly explains the procedure and gives a short summary of the patient's records.

A dialogue should be reached fast. A relationship of trust of the claimant in the SIP is necessary. Consequently, the SIP tries to explore the claimant's opinion on his situation. The SIP shows an attitude of empathy, respect and interest by continually asking questions and by taking subjective perceptions of the claimant into account. There is much room for the claimant to follow his line of thought and for the SIP to decide how he wants to conduct the interview, provided he pays attention to all five fields of the discussion. This leads to a light structuring of the interview. Precise questioning reveals the plausibility and consistency of the image that the claimant puts forward and how serious his incapacity is. The purpose of this method is to reach an understanding evaluation.

3 A physical examination is scheduled after the interview.

4 The SIP's final conclusion is stated clearly to the claimant, who is invited to react to that. The SIP presents his conclusion about limitations in functioning, with room to discuss remarks from the claimant. Then, the SIP explains the further procedure.

1. Health and disease (actual complaints, medical history, treatment and restrictions as experienced by claimant).
2. Work description (description and stressors).
3. Private situation (description and stressors).
4. Actual functioning (micro and meso, activities for the restoration of health and resumption of work).
5. Person (coping, locus of control etc.).
6. Physical examination.
7. Conclusion of the SIP, plan of action, if relevant, and plan of evaluation, if relevant.

Chapter 8:
General discussion

1 Background

The evaluation of work disability has a large individual and societal impact. Every year in the Netherlands tens of thousands of people are being evaluated on their claimed disability and granted or denied a benefit. The way these evaluations are carried out is a result of historical development and has changed considerably over the past two decades [1]. How can one be sure that laws on social insurance are being implemented correctly? As described in the Introduction to this thesis, in the Netherlands the quality of the evaluations that social insurance physicians (SIPs) perform is being debated. Individual SIPs have different ideas on what they are evaluating and how that evaluation can best be performed. Claimants and Institute for Employee Benefit Schemes ('Uitvoering Werknemersverzekeringen' or UWV) are implicated in this debate, too. A common understanding is needed in terms of what quality is and what quality can be expected and who is responsible for what aspect of quality control.
In this thesis the following questions are addressed:
1 What is the object of the evaluation of work disability?
2 What is to be understood by the quality of the evaluation of work disability?
3 How can the quality of evaluation of work disability be controlled?

Six studies have been completed:
1 Organisation of disability evaluation: a survey in fifteen countries
2 Medicolegal reasoning: a focus group and questionnaire study in four countries
3 Guidelines for evaluating work disability: an international survey
4 Evidence-based guidelines for evaluating disability: a comparison in two countries
5 Disability interview protocols in the Netherlands: a comparison of prescriptions and principles
6 Interviews for the evaluation of long-term incapacity for work: a study on adherence to protocols and principles

In this General Discussion attention is paid to the results of the different studies (par 2) and several methodological remarks made (par 3). Answers to the questions are formulated (par 4) and recommendations given (par 5).

2 Main findings

Chapter 2: Organisation and quality control of evaluation of disability in different countries.
In thirteen European countries, the United States, and the Russian Federation the organisation of disability evaluations was studied. Similarities and differences were found among legal criteria and in the way these criteria are used. In all countries studied, legal criteria for disability refer to 1) the claimant's ability (or inability) to perform work that can reasonably be expected from a worker in that profession, 2) health conditions that account for these abilities or inabilities, and 3) opportunities and obligations to receive treatment and return to work activities. These legal criteria are congruent with the criteria of the handicapped role. Legal criteria differ also; for instance, ability and inability are phrased as labour capacity or earning capacity and are rarely operationalised. The Dutch earning capacity is operationalised at the participation level of ICF while the British labour capacity is operationalised at ICF's activities' level. Among countries, differences exist in levels of disability, varying from one to seven. Finally, the waiting time before disability can be recognised is either fixed or variable and, in the first case, varies between six months and five years.
The organisation of the process of evaluation differs considerably among countries as well. The precise steps in the process and the connections with health care and labour market are organised quite differently. Legal criteria are operationalised in each country at the Institute of Social Insurance (ISI), as different types of output requirements for the social insurance physicians (SIPs): medical, functional, and rehabilitational. Most often, however, a country combines two or even three types.
The organisation of the evaluations can be seen as an interplay between actors that can be described using the extended script model. Actors in this organisation can be the SIP, the claimant, the ISI, an external supervisory body, the tribunals, the medical professional group, and the lawgiver. Explicit definitions of evaluation of quality were not found. An implicit definition at the legal level would be compliance with legal requirements; at the organisational level it would be compliance with organisational prescriptions, and at the professional level, compliance with professional standards.
Quality control is both indirect and direct. Indirect is, for example, the requirement by ISI that SIPs be qualified medical doctors. Indirect is the time and instruments doctors have for the evaluation and the specification of the required output. Direct quality control is seen in the inspection of case reports by staff doctors at the ISI. Two approaches emerge in quality control: individual SIPs who are monitored by the ISI and SIPs participating in medical committees who steer themselves.

Chapter3 Medico-legal reasoning: a focus group and questionnaire study
In four countries with different types of output requirements for their SIPs, the medico-legal reasoning of practising SIPs was studied in a case of an elderly construction worker with lower back pain. SIPs were asked to express and agree about the grounds they thought valid in their argumentation. This was first done in focus groups and secondly with a questionnaire, using grounds for argumentation as indicators of reasoning. SIPs in all countries studied proved to interpret disability in a way that meets legal criteria and the handicapped role. Added to that is a requirement of a fair trial. The handicapped role overlaps with the ICF: the situation of participation problems, activity limitations, and impairments, influenced by personal and environmental factors can be seen as an operationalisation of the health condition in the handicapped role. The grounds on the claimant's health condition and the grounds of medical evidence mostly are grounds for working capacity as well. Fair trial refers to plausibility, consistency (which can relate to ICF), exclusion of non-health related reasons for not working, and inclusion of the personal experience of the claimant.
The differences in output that were found in the study on organisation were not replicated in this study on medico-legal reasoning. The medico-legal reasoning does not simply follow the output requirements of the ISI but the more general handicapped role.
The grounds guide the translation of information into arguments about work disability. It is possible to make these grounds explicit.

Chapter 4 Guidelines for evaluation of work disability
From fourteen European countries the operationalisation of the evaluation of work disability and the use of guidelines were reported by central medical advisers and their staff. Five countries evaluate work disability in terms of all aspects of the handicapped role. In nine countries some aspects are not mentioned. Several countries report correspondence with ICF but nowhere the correspondence is made explicit. Official guidelines in social insurance medicine for evaluating work disability are found in Germany, Ireland, the Netherlands and Switzerland. Guidelines can be characterised as medical or procedural. Common topics of medical guidelines are the medical condition itself (including origin, risk factors, course, diagnostic procedures, treatment and ICD classification); return to work activities; and evaluation of work disability. This is an operationalisation of the handicapped role.
Common topics of the procedural guidelines are descriptions of the concepts of disease and incapacity to work as evaluated in social insurance, independence of SIP between claimant and ISI, qualification of the SIP, ways to determine the existence of disease, ways to determine functional capacity, and quality criteria of evaluations. These can all be seen as operationalisations of the concept of a fair trial.

Chapter 5 Evidence base of medical guidelines in evaluating disability
Two countries use medical guidelines: Germany uses six and the Netherlands sixteen. The quality of development of four pairs of guidelines on similar pathologies was studied using the AGREE instrument. All guidelines showed similar AGREE scores with only minor differences. Existing guidelines all meet the AGREE criteria of 'scope and purpose' and of 'clarity and presentation'. The procedures of looking for and incorporating evidence in the guidelines do not meet the AGREE criteria. The evidence is reported to be lacking for precise recommendations. The recommendations with regard to incapacity for work are expressed in non-specific, general terms. AGREE expects the guidelines to be drafted with involvement of all stakeholders and editorially independent. These requirements are only partly met. Client involvement is restricted and controversial. In Germany the guidelines are developed within the German Institute of Social Insurance (DRV). This reduces their independence.

Chapter 6 Disability evaluation interview protocols: comparison
The instrument that is used most often in daily practice is the interview with the claimant, which is susceptible to bias. One way of dealing with this bias is to use protocols. In the Netherlands three protocols have been developed to conduct the interviews in disability evaluation. These protocols were compared according to their similarities and differences through interviews with the authors of the protocols followed by a group discussion and comparison with existing scientific literature. The protocols all prescribe a semi-structured interview that varies in strictness. The topics that are prescribed vary in detail but all match the handicapped role and ICF. The procedural prescriptions aim at establishing a fair trial. The protocols are practice-based and have not been validated. The principles applied correspond with existing scientific findings but are not evidence-based.

Chapter 7 Disability evaluation interview protocols: application
The adherence of Dutch SIPs to interview protocols and their underlying principles was studied using a questionnaire among experienced Dutch SIPs. The results show a professional consensus about several basic assumptions. The principle of argumentative evaluation of disability and the principle of conducting a semi-structured interview are supported by over ninety percent of the respondents. Twelve interview topics are basically always addressed by over eighty percent of the respondents. All respondents used some form of protocol, either published or of their own making. Interestingly, no relationship was found between the SIPs' training and the use of a particular protocol. This consensus provides firm ground to develop further into principles of disability evaluation interviewing.

3 Methodological remarks

This thesis studies the evaluation of work disability, and this section discusses the strengths and weaknesses of this research.
Different articles focus on evaluating work disability and identifying its quality aspects, including medical, legal, and organisational aspects. Such a comprehensive approach is new. The Council of Europe [2] and Mabbet et.al. [3] report on the legal phrasing of criteria for disability but not on the way the criteria are applied. The OECD [4] has studied social insurance in many countries, but has focused on policies of granting benefits and promoting return to work, not on the evaluations.
Four studies make comparisons between countries; this is a new approach as well. By using the EUMASS network and other countries, a good variation in practices is reached. Comparing countries prevents a bias from one's national perspective but may limit the applicability of results in one particular country.
The research presented is mainly qualitative, establishing reliability through triangulation of data. For the questions at hand, and the state of science in the field, this seems appropriate. The findings are, on the whole, fairly consistent and can be related to existing literature. Many findings lend themselves to future quantitative research.
The question of what is being evaluated and how quality is defined and controlled in different countries is studied in legal criteria, in output requirements of the ISI and on the practice level of the SIPs. The legal criteria were found to be comparable and this seems reliable as reference was made to published texts. Earlier research [2,3] yielded more general but similar results. Organisational aspects like the definition of output and quality control were different among countries and this seems reliable, as reference was made to established policies and administrative prescriptions. Organisational aspects can probably be more differentiated as claims may be different. This thesis considers only straightforward first evaluation of work disability. In practice, organisational aspects may, within a country, vary from one region to another, a variation that was not studied. To have a valid picture of the full range of evaluations of work disability, these differences need to be described. Other studies with regard to the organisation of evaluation of work disability were not found and so cannot be compared. Practices were studied through questionnaires and focus groups but not examined in vivo; consequently, the results do not have a proven validity for the day to day work in disability evaluation. The terms that are used are not completely consistent among countries and thus the information may, at times, be flawed. This may be a matter of language but also of culture. This seems likely, for example, with regard to the question of what is evaluated in practice: some respondents distinguished between the health condition of a claimant and his functional capacity and scored both. Others probably did not make this distinction and scored only one of the two. Flaws may also have occurred when looking for procedural guidelines.

Some respondents did not perceive these as guidelines and did not report their existence. It seems unlikely that no prescriptions exist in these countries. A study on medico-legal reasoning (chapter 3) is interesting to do in more countries and based on more, and different, cases. All in all, the precise definition of the object of the evaluations in practice in different countries needs to be confirmed in further research.
The impact of treatment and coaching of sick leave for the evaluation of work disability has not received much attention in this thesis. Differences do exist and sick leave history is included in the evaluations (chapter 2) but this aspect needs more specific attention.
The different studies were performed over the past six years against a background of developing practice, society and science [5,6,7]. In the Netherlands disability evaluation now is done in a different legal scheme and under a different approach [8]. To the research questions of this thesis the changes probably do not influence the answers.
Explicit definitions of quality of work disability evaluation were not found in any country. This surprising result seems a reliable finding as it was explicitly addressed during all visits. Implicit definitions or common understandings probably do exist and need to be explicit in further research.
The policies to control quality were studied from the perspective of ISIs. One visible aspect of quality control is the use of guidelines. In this study several guidelines were identified and studied. It is likely that in practice more guidelines exist, albeit informally. It is also possible that local professionals have more mechanisms to ensure quality. This was not studied but it is worth the trouble to do so. The guidelines that were found were analysed from a medical point of view (handicapped role) and a procedural point of view (fair trial). The validity of these viewpoints for guidelines needs to be further confirmed as the material found was limited.

4 Answers to the questions:

This paragraph answers the research questions for the Netherlands and when appropriate from an international perspective.

Question 1: What is to be understood by evaluation of work disability?

At the legal level, as presented in chapter 2, the handicapped role [9] seems most fit to describe what is being evaluated. This is the case in all fifteen countries studied. At the legal level a fair trial is not mentioned but can be taken to be self evident.

At the organisational level, as presented in chapter 2, this is less clear. Different countries apply different output requirements that focus on aspects of the handicapped role: the medical condition, functional capacity, and the rehabilitation perspective. Clearly, different ISIs put different emphasis on aspects of the handicapped role. The handicapped role remains the core concept on the organisational level, however.

At the professional level, as presented in chapters 3, 5, 6, and 7, the handicapped role is most fit to describe what is being evaluated. In answering the questions in chapter 3, the respondents sometimes seem to equal the health condition with the (in) capacity for work. A requirement of fair trial is apparent in all four the countries studied [chapter 2].

In the situations studied in this thesis, evaluation of work disability is performed in a public arrangement. On legal, organisational, and professional levels, disability is defined differently. These findings are consistent with Lipsky [10] and Veen [11] who found that various levels in bureaucracies had different working definitions of their task. It is interesting that these differences exist and they must be an obstacle to a common policy of quality control.

Is a common answer possible? If we look at the results of the studies that are presented in chapters 2, 3, 4, 6, and 7, one answer is that an evaluation of work disability is *to conclude if the claimant meets the criteria of the handicapped role, based on an examination according to the practices that are considered appropriate in the arrangement at hand.*

This definition calls for operationalisation of two aspects: the handicapped role and practices that are considered appropriate.

The handicapped role refers to:

1 the claimant's restricted capacity to function in work
2 his damaged state of health as explanation of 1
3 his behaviour to recover and to take up work
4 possibilities to improve his health and functional capacity

Using the handicapped role agrees with existing opinions as expressed in Waddell & Aylward [12] and with the advice of the Health Council of the Netherlands [13]. The operationalisation of the object of disability evaluation by the Health Council [13] is similar to the handicapped role. The concept of role suggests more coherence in the evaluation than the seemingly independent tasks of evaluation as formulated by the Health Council of the Netherlands. The different aspects are not evaluated separately but can

be seen as conclusions that can be drawn from an evaluation of all four aspects. The Health Council of the Netherlands defines disability evaluation as the evaluation of 1) socio-medical history, 2) actual functional capacities, 3) the prognosis, and 4) actual treatment and socio-medical coaching.
This is not universally recognized in the Netherlands: current practice is to look at the evaluations as consisting of a main part, functional and earning capacity evaluation, and several more or less arbitrary parts [14]. In other studies, the evaluation of disability is presented as consisting only of evaluating functional capacity [15,16]. The Health Council of the Netherlands identifies evaluating actual functional capacities as the central task of the SIP. Conceptualising functional capacity as a separate characteristic of a claimant suggests an objective evaluation of functional capacity, which approach turns out to be difficult to accomplish [15,16,17]. Using ICF [18] as the main concept of disability in social insurance tends to support this approach. ICF was developed as a classification of consequences for chronic ill health, not as an instrument for disability evaluation in social insurance. In the scientific literature the ICF model is proposed as representing essential elements of disability [19]. In the studies presented here, ICF was not in official use in any legal scheme. ICF fits with what is evaluated but represents only part of it: the time perspective and the fair trial are not addressed.
The practices that are considered to be fit in the arrangement at hand refer to a fair trial. This aspect of the evaluation is less developed and more implicit than the handicapped role. Yet this consequence of the legal context in which the evaluations are performed is presumably universally present. The fair trial aspect is addressed in procedural guidelines that were found in four countries and possibly exist to some extent in all countries.
The operationalisation of disability evaluation by the Health Council of the Netherlands also indicates aspects of a fair trial formulated as quality conditions. In this study the fair trial seems, at the professional level, to have a more central role: without application of a fair trial an evaluation of work disability is not complete.

From the study on medico-legal reasoning (chapter 3) and in the guidelines from Germany and the Netherlands (chapter 4) and the disability interview protocols in the Netherlands (chapters 6 and 7), we can conclude that a fair trial refers at least to

1 Independence of the SIP towards claimant and ISI;
2 Ways to determine the existence of a plausible and consistent picture of impairments, disabilities, and handicaps;
3 Ways to determine functional capacity (parting from the claimant's opinion and verifying and completing this information).

Ad 1): The first requirement seems self-evident but is hardly operationalised. The script model of disability evaluation [20,21] illustrates that ways have to be found to establish a balance of interests. A way to do this on the individual level, as is demonstrated in chapter 6, is to use the introduction

part of the interview. On the group level of SIPs independence is shown by using guidelines for evaluation practice.
Ad 2): The second requirement fits with existing regulations but is well known to be difficult in practice. It rules out the requirement that disease exists only when a medical diagnosis has been proved, which can be particularly problematic in cases of unspecific diseases like chronic pain and fatigue. It is also challenged in the discussion between the doctors on the grounds that this is not a real medical, diagnostic, approach [22].
Ad 3): The third requirement fits existing regulations and opinions [8,23,24]. It shows that an evaluation of work disability is not a potentially objective measurement but rather a conclusion of a legal structure, based on conjecture and refutation. In Dutch literature this is called the argumentative evaluation of disability. This point of departure dictates a certain balance of power in the individual evaluation, the claimant having a substantial position.
To the extent that this answer of the first question is applicable in other countries needs more study. The present findings suggest that it is applicable. And no countries were found in which the answer does not apply. Different countries are in a different phase with regard to explicit definition of evaluation of work disability. Using the answer given here may help to speed up this development.

Question 2: What is quality of evaluation of work disability?
At the legal level, specific quality requirements were not found. Application of the law and proceeding with a fair trial are implicit quality requirements. At the organisational level, quality is organised rather than defined: specification of the output, evaluation by qualified personnel, and file inspection by colleagues are common measures. At the professional level, a definition of quality of evaluation was not found. In several countries guidelines have been produced. The medical and procedural guidelines represent professional agreement about the professional standards in the situations described in the guidelines. Implicitly, the professional definition of quality is therefore "performance according to professional standards." Direct determination and control of quality requires an operationalisation of quality. How can determination of quality of disability evaluation be realised?
Looking closer at the handicapped role, partly answers this question.

1 The capacity to function in work is a relational matter between a person's capacities and work demands. Measuring capacities in different situations yields different results [25]. A person's capacity to function in specific work depends partly on demands and circumstances of that particular work. It is difficult to imagine a measurement that includes all possible capacities under all possible circumstances. So called Functional Capacity Evaluations do give valuable information but cannot pretend to answer all questions [16]. Besides this technical problem, there is also a moral aspect: how much pain, sacrifice, and suffering

are reasonable to ask from a disabled person to endure to be (partly) economically independent?

2 The aspect of the damaged health condition requires this condition to be explanatory of insufficient capacity to work to meet the criteria for a benefit. Recent endeavours to support this aspect of the evaluations with medical guidelines invariably lead to the conclusion that a specific disease is likely to bring about certain restrictions in functioning, but mono-causal relations are the exception, not the rule. Many claimants are found to suffer from a combination of illnesses that have a combined influence on the claimant's capacity to function [13]. One Dutch procedural guideline prescribes that not the diagnosis but the consistent constellation of plausible manifestations of disease make up the requirement of a damaged health. This suggests much room for individual interpretation by SIPs, which is indeed found [15,26,27,28].
3 The aspect of behaviour towards recovery and resumption of work is partly empirical (what actions and interventions lead to recovery and resumption of work? What actions have been undertaken by this claimant?). Few cures or return to work programs are proven effective for the population that is evaluated for long-term incapacity for work [29]. Therefore, the question is: has he or she done and is he or she doing what can be reasonably expected, considering his or her situation and knowledge about effectiveness of certain interventions? So this evaluation is normative too.
4 The aspect about future possibilities is partly speculation about the answers of the earlier questions. The degree of certainty depends partly on these earlier questions supplemented by uncertainty about the future and by sound epidemiological evidence, if present.

It seems clear from these four aspects that the quality of the individual evaluations cannot simply be determined based on the quality of their output on the basis of measurable criteria that are independent of client and SIP. All four aspects are known to harbour dilemmas and no clear-cut answers are possible [14]. This explains why quality control in social insurance appears to be mainly indirect, as found by Meershoek et. al. [17]. This may seem unsatisfactory, but in the area of professional judgments is not at all exceptional [20]. Professional discretion is demanded in all situations where relevant grounds for the decision are determined by the situation itself, more than by rules or knowledge. Judges call this, in their verdicts, "the facts and circumstances of this particular case." This discretion can be handled in a sufficient fashion by ensuring the competence of the experts that perform the evaluations and by ensuring that they agree. This is partly common practice in social insurance already: we found that specific education of the SIPs is usual in many countries as is the inspection of doctors' reports by other doctors. A more direct way of enhancing SIP's agreement is to do evaluations in committees. This was found in Belgium and in former Eastern-European countries. This is an application of the

mechanism of Spearman and Brown that explains why a group opinion of experts is more likely to be correct than the experts' individual opinions [20].

Following this line of thought, quality of evaluations is found in the agreement of experts [30]. This is an intersubjective criterion instead of an external objective criterion. *Quality of evaluation of work disability is thus to be defined as experts' agreement on (aspects of) these evaluations.*
How can this quality be operationalised? And how free are the experts to follow their own opinions?
Operationalisation of the quality of evaluation of work disability can be found in how the evaluations are operationalised: to conclude if the claimant meets the *criteria of the handicapped role* based on an examination according to the *practices that are considered fit in the arrangement at hand*. Moving from the above answers, the operationalisation of quality includes:

1 SIPs agreement on the conclusions in terms of the handicapped role
2 SIPs agreement on the information gathering according to principles of fair trial

The study on Medico-legal Reasoning supports the idea that both principles guide SIPs in different countries. This conclusion is supported by the Guideline study and the studies on Disability Interview Protocols in chapters 4 and 6 respectively. As indicated earlier, the handicapped role seems to be broadly, although not universally, accepted. This is yet to be established of the fair trial.
One question is how far expert agreement is possible and necessary.
SIPs can agree on several levels:

- About principles and available knowledge in general. This agreement is found in education and by drafting evidence based guidelines. In this thesis, education was not studied. The procedural guidelines that describe the principles in the Netherlands are found to be relatively old and less coherent than the German and Swiss guidelines. The relevant knowledge as presented in the medical guidelines is limited and of general value.
- About application of principles and knowledge in individual cases. This agreement is found in sharing evaluations and case reports, and is done in training situations and peer tutorial sessions as used in the Netherlands. It is not known how effective these are in establishing agreement among SIPs. Their effect will be restricted to small groups and leave unanswered the question of the collective agreement.

To what extent this answer of the second question is applicable internationally also needs more study. The available evidence suggests that it is applicable. Countries like Germany and Switzerland have well developed procedural guidelines and the Netherlands and Germany have well developed medical guidelines. Other countries might fruitfully start from these guidelines to develop their own.

Question 3: How can quality of work disability evaluation be controlled?
Quality control consists of feedback loops from output or outcome of the evaluations of work disability to input or throughput [31]. At the legal level this means adapting the criteria according to outcome such as the disability statistics. This is done infrequently, although it does occur and then with great impact [6,7]. Quality control is usual at the organisational and the professional level. Feedback at the organisational level exists in many countries, notably in using the file inspection for feedback. The results are not published in any country studied. At the professional level, qualification is controlled mainly by continuous education.
From the previous reasoning it follows that an individual SIP can make an evaluation of good quality if he or she uses the expert consensus in every aspect of it.
Quality is primarily a matter of the professional group of SIPs. This description gives, however, not the complete image of quality control. The profession of Social Insurance Medicine is not a free market offer but is performed between parties described in the script model of disability evaluation. Consequently, SIPs need to decide about the quality of their work in a manner that convinces their environment: the contractor Institute for Employee Benefit Schemes (UWV), the claimant, and society as a whole. Claimant, UWV, and SIP are, according to the script of the evaluations, the first parties to contribute to quality control, each in their own fashion, individually and as a group. What do these parties do in the Netherlands?
The influence of the individual client is limited to the individual evaluation, which is not insignificant. The client is the object of the evaluation but also the subject. The better the claimant is instructed about the rules of the evaluation and the more aware he is of his situation, the better he will be able to claim a proper evaluation and influence the evaluation of his own disability.
Clients' organisations in the Netherlands are becoming more active in supporting their members and in participating in guidelines for SIPs. In the study about EBM guidelines [chapter 5] clients' organisations' involvement in medical guidelines was found to be in development but still restricted and controversial.

The individual SIP can realise quality by gathering information according to professional rules of evaluation and by linking this information to professionally accepted grounds into arguments that conclude about the client's handicapped role. Like many professionals, the SIP stands between managerial instructions that often are considered mechanic [32,33] and full autonomy to decide, which is considered to lend itself to the use of strictly personal convictions [17, 30]. A balanced position seems to be to rely on professional standards that satisfy the administration and that leave room for tailor-made evaluations [34,35]. This requires that the SIP participates in permanent professional education and that he or she has the attitude and the situation to work according to professional standards.

For this, the SIP needs to know the existing professional consensus. The guidelines describe the consensus to some extent. Medical guidelines define criteria and grounds with regard to specific diseases. However, the medical guidelines are not strongly evidence-based and give only general answers to the individual questions that the SIPs face. The Dutch procedural guidelines provide some answers with regard to rules, but these answers are relatively old and lack the coherence of the German and Swiss guidelines. About the social norms, no explicit sources are available. SIPs can be more explicit by identifying the grounds they base their arguments on. These grounds can be characterised in different ways; one is according to their source: knowledge (preferably scientifically proven), rules (preferably legally based), and norms (preferably socially shared). See Figure 1:

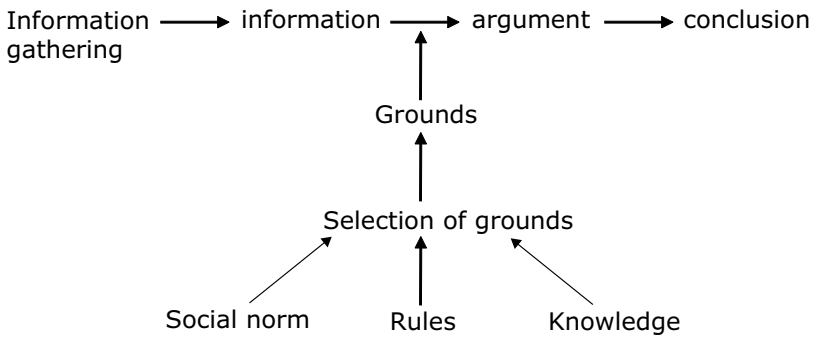

Figure 1: Argumentation and sources of grounds.

At the group level, it is new that the Dutch SIPs define their professional standard in guidelines. This traditionally is the domain of the UWV as is the case in other countries [chapter 2]. In the study on organisation of disability evaluation, it was found that much room existed for active control of quality, especially in terms of professional standards. Since 2003, sixteen medical guidelines have been drafted by the Health Council of the Netherlands and by the Dutch society of Social Insurance Physicians (NVVG). These guidelines support the evaluations but any accounting of how these guidelines are applied in daily practice is missing, except for the case report inspection by UWV staff. The results of these inspections are not published. Even more recent is the advice of the Health Council of the Netherlands that the professional group develops medical case law *("mediprudentie")*, a collection of well documented and professionally commented case descriptions. These case descriptions are to demonstrate the professional consensus on a case by case base. Recently this approach was tested [36] and it is being implemented now.
UWV, the contractor of the evaluations, can facilitate and check the application of professional standards of the SIPs, which the SIPs see as

problematic [14, 15, 37]. The study on organisation of disability evaluation [chapter 2] dates from 2003 and is partly outdated for Dutch practice in 2009. Since 2003 guidelines have been implemented at UWV and peer tutorial sessions have been extended to draft illustrative cases (precursors of *mediprudentie*, see above). A consistent policy of monitoring and controlling quality has, however, not been officially published.
Internationally this answer again seems applicable. However, the balance between ISI and the professional group appears to differ among countries. Applying this approach calls for tailor made solutions.

The parties most concerned with quality control of evaluation of work disability are ISI, SIPs and their organisations, and claimants and their organisations. Their mutual influencing can be summarised as in Figure 2:

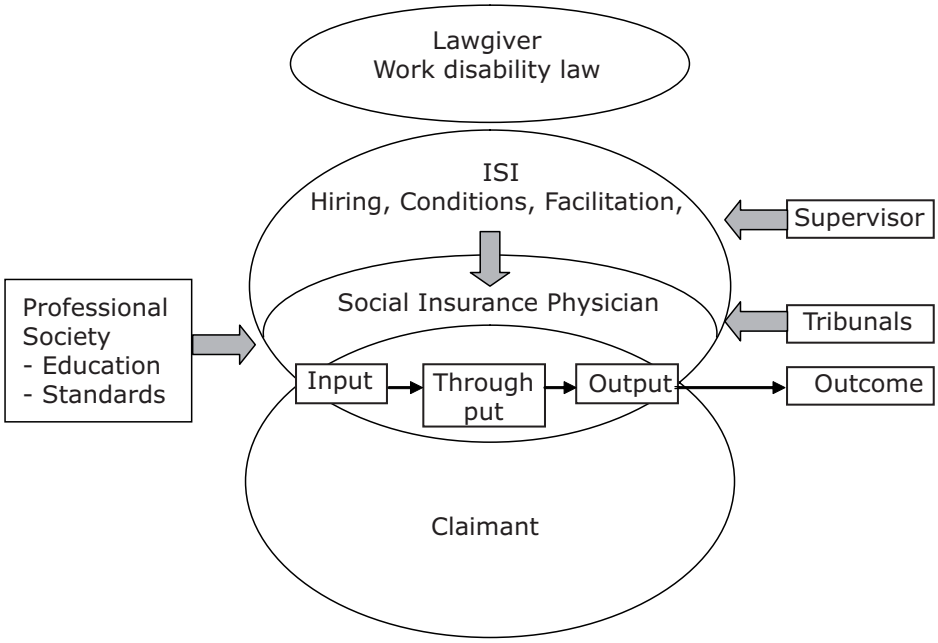

Figure 2: Parties in quality control of evaluation of work disability.

5 Recommendations:

The conclusions suggest certain actions that can be taken to improve quality in evaluating work disability in the Netherlands. Quality needs to be realised at the practice level, so the recommendations will focus on this level first. Next, we will consider quality control at management and other levels. When applicable, recommendations for other countries will be provided.

Practice level

The main challenge for quality improvement in the Netherlands is for practicing SIPs. They are unsure about the quality of their work and have to resolve these uncertainties by using practices and grounds that represent expert consensus and scientific evidence. Their job is to perform the disability evaluations according to professional guidelines and standards that are constantly evolving.

Making the grounds for disability evaluation explicit in the reports of individual evaluations will enhance the quality of these medical reports, as the grounds used can be evaluated. Developing consensus about these grounds would be a crucial step forward, as consensus will improve professional practice and contribute to the transparency and legitimacy of the disability pension schemes that are open to constant criticism about their capacity to select the right people for a disability benefit.

In terms of the practice of evaluating incapacity for work, both medical and procedural guidelines are important instruments to improve and control quality. It can be expected that professionals' guidelines will endorse the legal security of claimants as well. For these reasons, the development, testing, and implementation of guidelines deserve support.

Consensus about the interview routine can be used to further develop professional consensus on SIPs' attitude and on the structure of the interviews. Without striving for a detailed and universally applicable protocol, relevant topics can be selected that are well circumscribed and based on consensus and evidence about what constitutes disability.

If such measures are developed, implemented, and controlled, the quality of the evaluations will improve in terms of transparency and reproducibility. The evaluations will become more comparable and more accessible to scientific research on behaviour of both the SIP and the claimant. The proposed innovations will enable claimants to better prepare themselves for the evaluations, which will make their position more equal to that of the SIP, in favour of the satisfaction of the claimants.

The professional group

It is their professional society (NVVG) that faces the challenge of uniting the SIPs and of developing professional practice guidelines and the medical case law. An essential first step is that the professional group agrees that the evaluation of work disability is not limited to the evaluation of

functional capacity. Evaluation of work disability needs to be based on operationalisations of the handicapped role and on the principles of a fair trial. Another essential step is to define when an evaluation can be said to meet quality standards, including input, process, and output aspects of the evaluation. The development of medical case law will result in many examples.

Over the past years many initiatives in professional development have been implemented. Further action is necessary to consolidate the scientific tradition in social insurance medicine as a stable infrastructure is not established yet. Three developments are particularly promising for the near future: harmonising the procedural guidelines, revising medical guidelines according to AGREE criteria, and developing medical case law.

Procedural guidelines have been developed in the past but in a piecemeal fashion. A guideline on the evaluation of clients' behaviour towards recovery and resumption of work is underway. Procedural guidelines would profit from more cohesion to improve their applicability and transparency. An option is to produce one comprehensive new procedural guideline. The guidelines need to be updated anyway.

Medical guidelines will need updating in the near future as well. Applying AGREE criteria more rigorously will help in their development and will indicate the lack of evidence and promote targeted research. The basic conditions for optimal guideline development can be met already.

Medical case law (mediprudentie) may be one of the core instruments in the coming years as it is a promising way to demonstrate the professional standards on a case by case basis. Medical case law indicates how principles and knowledge can be applied in individual cases.

Consensus about the essentials of a disability interview will substantially contribute to valid evaluations and good documentation. This consensus can provide a starting point to develop a guideline that can be implemented in practice and evaluated. For implementation, more than a single training will be needed. Some form of supervision and support will be necessary to maintain and improve the application. The everyday practice of a growing diversity of claimants calls for research and development in communication for SIPs. Companies, medical specialists, and occupational physicians will need to be trained for their new roles as well.

Internationally, it seems possible and worthwhile to harmonise the efforts of the professionals. As seen in the study on medico-legal reasoning, the handicapped role and the fair trial turns out to be pivotal for the evaluation of work disability per se. Possibly this applies further than work disability, to participation problems in general. Studies on medico-legal reasoning, but also the making of case descriptions, should be tried internationally as well, perhaps in this order so as to find any incompatibilities in professional grounds. The EUMASS network can facilitate this for European countries.

Institute of Social Insurance (ISI)
All actions where professionals take initiatives, as mentioned above, need to be performed in collaboration with the ISI and with claimants' representatives. The Dutch ISI, UWV, is responsible for both the administrative and medical quality of the deployment of the law, notably the 'Work and Income, according to Labour Capacity Act' (WIA). Overlap in responsibilities of ISI and the professional group easily leads to tensions between UWV, SIPs, and professional organisations. Professional quality control will benefit from a supportive and challenging attitude of UWV toward the professionals. One innovation can be that the SIPs report the functional capacity of claimants and other aspects of the handicapped role. The application of a fair trial needs to be elaborated and implemented. Considerations of efficiency are relevant for both ISI and professionals and need to be weighed against professional standards. A transparent trade-off will help make tensions clear and manageable.

Claimants
Claimants as individuals are not easily in a position to demand and develop quality evaluations as they are object in the evaluations. But claimants are in a position to identify problems in daily practice, and their representatives can contribute to the development of professional guidelines. Recent experiences in the Netherlands with guideline development suggest that their input can be very valuable. This deserves support as the representatives of the clients can be the SIPs' allies in the quest for quality.
In other countries too, claimants can be involved in developing guidelines, if it is possible to find clients' representatives who can share the basic tenets of disability evaluation.

Supervisor
The Ministry of Social Affairs and Employment monitors the performance of ISI in applying the legislation on work disability. Consequently, they are in a position to demand, facilitate, and verify that SIPs, ISI, and claimant organisations cooperate to promote the quality of evaluations. The model of the script of evaluation suggests that a countervailing power is needed. This prevents a one-sided demand on efficiency by the ISI opposing professionals' guidelines that would be detrimental to the quality and effectiveness of the disability evaluation. The Ministry, representing the public interest, is in a position to provide checks and balances, and has the opportunity to involve the Health Department to warrant cooperation between the worlds of health and labour [38].

Future Research
All recommendations provided will need to be supported by scientific evidence and proper scientific evaluation. Studies are needed to verify if guidelines are helpful to establish a good quality evaluation. Various questions need to be answered: How can medical guidelines be developed that are more

firmly evidence-based? How can implementation be stimulated in practice? What is the effect of procedural guidelines in the effort to improve and control quality? Does medical case law enhance consensus between SIPs about the correct grounds to use and about proper information gathering and use of evidence? Answering these questions is a matter of independent research. Practical instruments to establish agreement of experts need to be developed, which can serve as gold standards to test the effectiveness of the suggestions made.

The guidelines would profit from considering the handicapped role as the object of evaluation and from more specific recommendations. For this, and to support individual evaluations, epidemiological research is needed on the functional capacity that people with disabilities have and how they deal with these capacities in their daily lives. With regard to less common diseases, the evidence on the prevalence and incidence of impairments will always be limited. With the introduction of new therapies too, much will be unknown about the impact on people's work capacity. An alternate, more generic approach for chronic conditions is to use existing evidence about the generic effects of chronic medical conditions and diseases on working capacity and participation opportunities. Epidemiological studies departing from this point of view can be more effective than starting new studies for every disease separately.

The difficulties between professionals and their employers are by no means specific for SIPs. Both employers and professionals seek an optimum way of achieving excellence within boundaries of efficiency, which is an interesting field of research that can help solve these problems. SIPs provide good case examples as they form large populations that are relatively homogeneous in their tasks and are, on that task, comparable among countries.

So far, in discussing the quality of the evaluations, emphasis has been on structure, process, and output aspects such as education of SIPs, quality of guidelines, good application of the guidelines, and good communication during the evaluation following a well developed interview protocol. In the end, evaluations result in people with or without a benefit and with or without work or accommodations. It would be interesting to monitor the fate of claimants on the labour market after their evaluation. This would provide important feedback to the SIPs about the opportunities people with disabilities in reality have on the labour market.

Most of these recommendations for research are relevant in other countries too. It is interesting to note that in disability evaluation so many similarities are found among countries. That opens fascinating perspectives for the international development of the social insurance medicine discipline, in education and in research. The adoption of guidelines and case law elsewhere would show the effect of different conditions in different countries. The same goes for epidemiologic study of populations of claimants and people who receive benefits. Another interesting field is the question of supposed incompatibility between evaluating clients and trying to help the evaluated clients. This is done differently in different countries

and international comparison would help to find opportunities that are not found yet. Finally, professionals working in administrative organisations are found in all countries studied in this thesis. The aspect of quality control in such circumstances also deserves international comparison.

Future quality

Evaluation of work disability is a most relevant activity from both an individual and a social perspective. It is underrated if its societal impact is compared to the efforts to study and improve daily practice. Many causes can be identified for this lack of scientific interest and the lack of profiling and professional pride of the social insurance physicians are among them. In the Netherlands over the past few years, quite a few initiatives have been taken to scientifically endorse the evaluation of work disability. The findings in this thesis show the need to continue. If the recommendations are brought into practice and are effective, professionals can enhance the quality of their work. The legitimacy of the system and the sense of an equitable treatment with the claimants will be supported by this. That is an opportunity not to lose.

References

1 Boer WEL de, Houwaart, ES. (eds) **Ability evaluated: work disability evaluation in the Netherlands 1901-2005**. *[in Dutch]* Utrecht NVVG 2006

2 Council of Europe. **Assessing disability in Europe**. Strasbourg, CoE 2002

3 Mabbet D, Bolderson H, Hvinden B. **Definitions of disability in Europe: a comparative analysis**. Uxbridge Brunel University 2002

4 Organisation for economic cooperation and development (OECD). **Transforming disability into ability**. Paris OECD 2003

5 Organisation for economic cooperation and development (OECD). **Sickness, Disability and Work (Vol. 1): Norway, Poland and Switzerland**. Paris OECD 2006

6 Organisation for economic cooperation and development (OECD). **Sickness, Disability and Work (Vol. 2): Australia, Luxembourg, Spain and the United Kingdom**. Paris OECD 2007

7 Organisation for economic cooperation and development (OECD). **Sickness, Disability and Work (Vol. 3): Denmark, Finland, Ireland and the Netherlands**. Paris OECD 2008

8 Health Council of the Netherlands. **Evaluation, treatment and coaching**. *[in Dutch]* 2005/10 Den Haag Gezondheidsraad 2005

9 Gordon G. **Role theory and illness**. New Haven College and University Press, 1966

10 Lipsky M. **Street level bureaucracy**. New York Russell Sage, 1980

11 Veen RJ van der. **Social restrictions of politics**. *[in Dutch]* Leiden Stenfert Kroese 1990

12 Waddell G. and Aylward M. **The scientific and conceptual basis of incapacity benefits**. Norwich The Stationary Officer 2005

13 Health Council of the Netherlands. **General introduction to the guidelines for social Insurance physicians**. *[in Dutch]* 2006/22 Den Haag Gezondheidsraad 2007

14 Boer, WEL de, Steenbeek, R. **Report on problematic situations in disability evaluation**. *[in Dutch]* Utrecht NVVG 2006

15 Bont AA, Berendsen L, Boonk MPA, Brink JC van den. **At the doctor's office**. *[in Dutch]* Zoetermeer CTSV 2000

16 Wind. H. **Assessment of physical workability: the utility of Functional Capacity Evaluation for insurance physicians**. Amsterdam PhD Thesis Amsterdam University 2007

17 Meershoek A, Krumeich A, Vos R. **Judging without criteria? Sickness certification in Dutch disability schemes.** *Soc Hlth Illness* 2007 **29**: 497-514

18 World Health Organisation (WHO) **International Classification of Functioning and Health**. Geneva WHO 2001

19 Kierkegaard M, Harms-Ringdahl K, Holmqvist LW, Tollbäck A. **Perceived functioning and disability in adults with myotonic dystrophy type 1: A survey according to the International Classification of Functioning, Disability and Health** *Journal of rehabilitation medicine* 2009 **41**: 512-520

20 Hofstee WKB. **Principles of evaluation. Methods and ethics of selection, exams and evaluation**. *[in Dutch]* Lisse Swets & Zeitlinger 1999

21 Boer WEL de, Hazelzet AM, Gerven JHAM van. **Manual of evaluations of disability**. *[in Dutch]* Hoofddorp TNO 2002

22 Wolfe F. **The fibromyalgia wars**. *Journal of Rheumatology* 2009 **36**: 671-678
23 **Law on work and income according to working capacity (WIA)** *[in Dutch]* http://lexius.nl/wet-werk-en-inkomen-naar-arbeidsvermogen visited 4-9-2009
24 Council for Health Research. **Advice on social Insurance medicine**. *[in Dutch]* Den Haag, RGO 2004
25 Brouwer S, Dijkstra PU, Stewart RE, Goeken LN, Groothoff JW, Geertzen JH. **Comparing self-report, clinical examination and functional testing in the assessment of work-related limitations in patients with chronic low back pain**. *Disab & Rehab* 2005 **27**: 999-1005
26 Kerstholt J, Boer WEL de, Jansen NJ. **Disability assessments: effects of response mode and experience**. *Disab & Rehab* 2006 **28**: 111-115
27 Slebus FG, Sluiter JK, Kuijer PP, Willems JH, Frings-Dresen MH. **Work ability in sick-listed patients with major depressive disorder**. *Occupational Medicine* 2008 **58**: 475-479
28 Spanjer, J., Krol, B, Brouwer, S., Groothoff JW. **Interrater reliability in disability assessment based on a semi-structured interview report**. *Disab & Rehab* 2008 **30**: 1885-1890
29 www.bimra.nl *[in Dutch]* visited 4-9-2009
30 Hofstee, WKB. **Promoting intersubjectivity: a recursive-betting model of evaluative judgments**. *Netherlands Journal of Psychology* 2009 **65**: 2-9
31 Imaki, M. **Kaizen, the key to Japan's competitive success** New York Random House 1993
32 Brink, G van den, Jansen T, Pessers D (eds) **Professional pain. Why the Netherlands don't work properly**. *[in Dutch]* Amsterdam Uitgeverij Boom 2005
33 Weggeman, M. **Managing professionals: just don't!** *[in Dutch]* Schiedam Scriptum 2007
34 Freidson, E. **Professionalism, the third logic**. Cambridge Cambridge Polity Press 2001
35 Scientific council for government policy. **Proofs of providing proper services**. *[in Dutch]* Den Haag WRR 2004
36 Committee for preparing medical case law. **Medical case law at the start**. *[in Dutch]* Utrecht NVVG 2009
37 Choy, YW. **Report about the consultation of social insurance physicians and labour experts by the department of social- medical affairs**. *[in Dutch, not officially published]* Amsterdam UWV 2005
38 **Blue print of Participation in Medical Guidelines** *[in Dutch]* Utrecht VGI 2007

Appendix: Evaluation of work disability in the Netherlands (2009)

Normally the applications for a work disability benefit are treated as follows:

Legal conditions:
To qualify for a disability benefit, a salaried worker has, due to disease or handicap, to be unable to earn more than 35 percent of what a comparable healthy person earns in customary work. In this definition it is the loss of earning capacity that is insured, not the health problem itself.
A salaried worker can file a claim after a period of sick leave from his or her former function of two years. The claim has to be filed at the UWV, the Institute for Employees Benefit Schemes. UWV has the work disability evaluated by a social insurance physician (SIP) and a labour expert.

Procedure:
The SIP performs a medical examination and estimates the claimant's current ability to function. The SIP then drafts a general medical report. The SIP may consider the claimant fully and permanently disabled for work and report so. Otherwise, the SIP completes a functional capacity report (FML), which contains 56 items in six categories:
Personal Functioning: focus of attention, remembering, speed of action, etc.
Social Functioning: seeing, hearing, dealing with emotions of others, conflicts, etc.
Adaptation to Physical Environment : heat, cold, draught, allergies, etc.
Dynamic Activities: the use of hand and fingers, walking , squatting, climbing stairs, etc.
Static Positions: sitting, standing, ability to work above the head, etc.
Working Hours per day or per week.
The SIP selects the relevant categories and ticks the relevant items as capable, partly capable, or incapable. The labour expert uses this information to identify existing jobs that the claimant can theoretically work in. The FML is designed so that it can be used as input in a computer-based matching programme. In this programme some 7500 jobs nationwide are described, with their requirements and earnings. This information is updated on a regular basis by job analysts employed by UWV. The programme provides examples of jobs that the claimant can theoretically do, based on the functional capacities as identified by the SIP. The labour expert then selects those jobs that seem feasible in the case of the claimant. If this selection leaves sufficient jobs that the claimant can theoretically do, the average earnings of these are calculated and compared to the claimant's earlier earnings. This may show a percentage of loss of earning capacity that is the basis for a benefit if the loss is over 35 percent. The benefit is proportional to the loss of earning capacity.

Benefit:
Benefits are time limited and claimants are reassessed regularly for continuation of the benefit. If the SIP considers the claimant fully and permanently disabled for work, the claimant is granted a full benefit. If the labour expert finds a percentage of loss of earning capacity over 35 percent, that is the basis for a partial benefit. The benefit is proportional to the loss of earning capacity. The claimant is expected to look for work that fits his or her capacities and he or she is periodically reassessed.

Quality control:
The evaluation of work disability is subject to several measures to control quality:

1 The SIPs and labour experts receive training within UWV and in official post-graduate education (SIP) and high school (labour expert). They are expected to continue their professional education throughout their professional lives.
2 SIPs and labour experts are expected to participate in peer tutoring sessions where they discuss professional aspects of their daily work and functioning.
3 Guidelines have been drafted for the work of SIP and labour expert. These are to support their professional decision making. These guidelines can be found at www.cba.uwv.nl.
4 Reports of SIPs and labour experts are sampled by UWV staff for review of the case handling. SIPs and labour experts receive feedback on the result of this.
5 Polls among clients of UWV in which their opinion of the service delivery of UWV is asked.

Summary

This PhD thesis addresses the evaluation of work disability as practiced in the Netherlands under the Work and Income according to Labour Capacity act (WIA) and the Act on Work Disability for people who became handicapped in their youth (WAJONG), both successors of the Dutch WAO and AAW acts. Chapter 1 describes how these evaluations are performed and why they are worth studying. WIA and WAJONG are implemented by the Institute for Employee Benefit Schemes (UWV). Social insurance physicians (SIPs) and labour experts work for UWV in performing the disability evaluations and advising the administration on whether to pay (a full or partial) benefits and advising claimants about their possibilities for work. Such evaluations greatly impact individual claimants whose immediate social and economic positions depend on these evaluations. Taken together, the evaluations and benefits also have a large societal impact.
However, doubt exists about the quality of the evaluations among both claimants and SIPs. Considering the many reforms of scheme and organisation of the past decades, this doubt is understandable. The evaluations, largely based on interviews with claimants, are susceptible to subjective influences. However, it is surprising that so little is known about the evaluations and their quality. For this reason, the following questions are addressed:

1. What is the object of the evaluations of long-term work disability?
2. What is the quality of these evaluations?
3. How can the quality of the evaluations of long-term work disability be controlled?

The question of what is actually being assessed can be approached in different ways. SIPs agree that they evaluate working capacity, but they disagree on the precise definition of working capacity and whether that is all they evaluate. This uncertainty is quite reasonable, as the legal criterion of work disability is rather abstract: *"being incapable, as a consequence of direct and medically certifiable illness or defect, to earn at least 65% of what a comparable healthy person can earn in usual work"*.

One approach to this problem can be found in the WHO's model of the International Classification of Functioning and Health (ICF), a model of the consequences of disease. In the ICF, illness, impairments, and restrictions of activities and participation are classified in relation to personal and environmental factors. Apart from the legal criterion of work disability, other legal provisions are relevant for the evaluations. These provisions have given rise to professional guidelines from the Tica of the 1990s. The Health Council of the Netherlands has, by another approach, identified four aspects in evaluating work disability: the socio-medical history, the actual functional capacity, the actual treatment and coaching of the claimant, and the prognosis. This fits with the legislation and the opinion

that work disability is a process that is considered from the perspective of a policy goal of promoting participation in work. Yet another approach is the handicapped role as described by G. Gordon. This is an extension of the sick role as stated by T. Parsons. This approach puts the action of the individual at a central position. A person who cannot work normally because of illness or defect can claim a partial exemption and compensation but is also obliged to strive for recovery and to account for his/her situation.
The question about the quality of the evaluation of something as unspecific and immeasurable as work disability has hardly been studied. Usually, the quality requirements are left to the evaluating experts themselves, which is the case in the field of work disability; however, thus far experts have not explained their concept of quality and whether that concept works in practice.

The problem of quality control brings up the question of who is concerned with the evaluations. In a script model of evaluation, Hofstee describes in universal terms the parties concerned with an individual evaluation. For work disability, the script can be operationalised and the concerned parties, apart from the claimant and the assessor, are at minimum the disability agency (UWV) and the professional group of evaluators.

Using these approaches, six studies have been performed, both in the Netherlands and elsewhere, and are described in this thesis.
Chapter 2 describes how the work disability evaluation is organised and how its quality is controlled in fifteen countries. This analysis was performed using a literature study, a questionnaire, and visits. Legal criteria turn out to be quite comparable among countries, but there are considerable differences in how their application is organised. Quality control is, in general, indirect and implicit.
Chapter 3 discusses the reasoning of SIPs in the case of an elderly construction worker who claims work disability. This study was done using focus groups and a written case, followed by a questionnaire in larger groups of SIPs in four countries. The grounds that the SIPs use proved to fit with the handicapped role and a requirement for a fair trial. There was good agreement about how to assess functional capacity but much less agreement on the question of what can be asked from a claimant before permanent disability can be accepted.
Chapter 4 describes the operationalisation of work disability of SIPs in thirteen countries and the guidelines that exist for their evaluations. These topics were studied using a questionnaire completed by central medical advisers of institutions of social insurance and interviews in the countries where the guidelines were said to be used. Four countries use officially prescribed guidelines that are either medical or procedural. The medical guidelines support the evaluation of the handicapped role; the procedural guidelines support the fair trial.
Chapter 5 tests the scientific development of medical guidelines in Germany

and the Netherlands using the AGREE instrument, which is a tool that can be used to assess the quality of guideline development. Thus far, the AGREE instrument has been used mainly in the clinical domain, but it is applicable for SIP medical guidelines as well. The guidelines that were tested are sufficient in several aspects, but not in terms of stakeholder involvement, rigour of development, and editorial independence.
Chapter 6 describes the three Dutch protocols for disability evaluation interviews. This work is based on literature about the protocols, interviews with those who drafted the protocols, and literature discussing the assumptions that the protocols are based on. The protocols are compared with regard to what they suggest should be discussed during the interview and how the interview should be conducted. In both aspects, the three protocols prove very similar, but they show differences in the prescribed strictness of application and the need to activate the claimant during the interview, apart from assessing his/her functional capacity.
Chapter 7 describes the adherence of Dutch SIPs to the interview protocols, using a questionnaire study of 150 physicians. All do use a protocol of some form, either published or of their own making. There is no significant relation between being trained and following a particular protocol. Over eighty percent of the respondents agreed on the topics that constitute the handicapped role.
Chapter 8 summarises the results and comments on the methodology used. The actual practice of disability evaluation is not studied in this thesis. In addition, attention is focused on ordinary first assessments and not on re-assessments or especially complex cases. Procedural (and to a lesser degree medical) guidelines are probably used more often than found in these studies.

The conclusion is that evaluations of work disability assess how well the claimant fulfils the handicapped role. This complies with legal and policy goals as well as the medical views, for instance, of the SIPs. Furthermore, striving for a fair trial is an important criterion of the evaluations and complies with application of a law. The definition of quality and its operationalisation is, first, up to the experts. In several countries, experts have been active in quality control. It is obvious that the definition of quality is not solely determined by experts: their contractors and the claimants can influence this measure. Much can be gained by quality control. One recommendation is to challenge and facilitate the experts to assume more responsibility. The experts can draft and improve medical and procedural guidelines, and further develop effective interviewing practices and medical case law. This all depends on a proper scientific grounding, as the professional debate has been fraught with considerations rather than empirical findings.

Samenvatting

Dit proefschrift gaat over de beoordeling van langdurige arbeidsongeschiktheid in het kader van de WIA en de Wajong, de opvolgers van de WAO en de AAW. In hoofdstuk 1 wordt beschreven hoe die beoordelingen plaatsvinden en waarom die het onderzoeken waard zijn. De wetten worden uitgevoerd door het UWV en door bij of voor UWV werkende verzekeringsartsen en arbeidsdeskundigen die de zogeheten claimbeoordelingen doen. Die beoordelingen leiden tot adviezen aan UWV om wel of niet, volledig of deels een uitkering toe te kennen, en tot adviezen aan de cliënt, de werknemer die een claim indient, over mogelijkheden om te werken. Individueel heeft de beoordeling veel betekenis voor de cliënt wat betreft zijn sociaal economische positie die door de beoordeling kan worden bepaald. Alle adviezen en uitkeringen als geheel hebben een grote maatschappelijke impact.
De kwaliteit van de beoordelingen is onderwerp van discussie bij zowel cliënten als beoordelaars, in dit proefschrift de artsen. Die discussie is niet verwonderlijk, mede gezien de vele veranderingen in regelgeving en organisatie die de afgelopen decennia hebben plaatsgevonden. Ook zijn de beoordelingen, die voor een belangrijk deel zijn gebaseerd op gesprekken met de cliënten, kwetsbaar voor subjectieve invloeden. Het is opvallend dat er zo weinig bekend is over de vorm en inhoud van de beoordelingen en over de daarbij te leveren kwaliteit.
Om die reden wordt in dit proefschrift gezocht naar antwoorden op de volgende drie vragen:

1 Wat wordt beoordeeld bij de beoordeling van langdurige arbeidsongeschiktheid?
2 Wat is kwaliteit van die beoordelingen?
3 Hoe kan de kwaliteit van de beoordelingen van langdurige arbeidsongeschiktheid worden geborgd?

De vraag naar wat wordt beoordeeld kan vanuit verschillende gezichtspunten benaderd worden. Verzekeringsartsen zijn het met elkaar eens dat ze de aanwezigheid van mogelijkheden om te functioneren beoordelen maar minder eenduidig is de overeenstemming over wat dat is en over verwante aspecten die wel of niet worden beoordeeld. De wet zelf geeft een vrij abstract criterium voor wat arbeidsongeschiktheid is, namelijk het als *rechtstreeks en medisch objectief vaststelbaar gevolg van ziekte of gebrek niet in staat zijn om in gangbare arbeid meer te verdienen dan 65% van wat een vergelijkbare gezonde persoon gewoonlijk verdient.*
Over de relatie tussen ziekte en functioneren en maatschappelijk participeren heeft de WHO een model opgesteld dat de gevolgen van ziekte classificeerbaar maakt. In dat gevolgenmodel worden ziekte, stoornissen, beperkingen in activiteiten en participatie benoemd, in samenhang met persoonlijke factoren en omgevingsfactoren.
De wet geeft, naast het genoemde criterium voor arbeidsongeschiktheid,

meer bepalingen die voor de beoordeling relevant zijn en waar ook professionele richtlijnen voor zijn gemaakt door het Tica in de jaren 90 van de vorige eeuw. De Gezondheidsraad heeft voor het ontwikkelen van de protocollen voor de verzekeringsgeneeskundige beoordeling aangenomen dat er vier taken van beoordeling zijn, namelijk het beoordelen van de sociaal-medische voorgeschiedenis, van de actuele functionele mogelijkheden, van de behandeling en begeleiding en van de prognose. Deze indeling past goed bij de wetgeving en bij de opvatting dat arbeidsongeschiktheid een moment is in een proces en dat dat proces bezien dient te worden vanuit een beleidsdoelstelling van bevordering van participatie. Een wat oudere en minder bekende benadering die relevant kan zijn is die van de gebrekkigenrol, een concept van Gordon uit 1968 dat voortbouwt op het bekende concept van de ziekenrol van Parsons uit 1953. In de rolbeschrijving van Gordon staat de actie van het individu centraal. Diegene die vanwege ziekte of gebrek niet goed kan meekomen heeft aanspraak op gedeeltelijke vrijstelling en tegemoetkoming maar deze heeft ook een verplichting om zich naar vermogen in te spannen en aan te passen en om daarover verantwoording af te leggen.
De vraag naar de kwaliteit van een beoordeling van een weinig specifiek en lastig meetbaar iets als arbeids(on)geschiktheid, is tot op heden weinig onderzocht. Het is in veel gevallen gebruikelijk om kwaliteitseisen voor professionele arbeid over te laten aan de experts die die arbeid uitvoeren. Dat is ook in dit domein gebruikelijk, maar tot nog toe hebben de experts niet voldoende uitgelegd wat zij onder die kwaliteit verstaan en hoe we kunnen weten dat die goed is in de praktijk.
De vraag naar de borging van de kwaliteit leidt automatisch naar de vraag wie betrokken zijn bij die borging. Dat zijn diverse partijen die door Hofstee 1999 in universele termen zijn beschreven in een script. Voor de beoordelingen van arbeidsgeschiktheid kan dat script worden toegespitst. Centrale partijen zijn in ieder geval, naast de cliënt en de beoordelaar, het UWV en de beroepsgroepen van beoordelaars. Overheid en sociale partners en anderen spelen in de dagelijkse praktijk een rol op de achtergrond.
Gebaseerd op deze benaderingen zijn zes onderzoeken gedaan, deels in Nederland, deels ook in andere landen, die in verschillende hoofdstukken zijn beschreven. In hoofdstuk 2 wordt weergegeven hoe de beoordeling van arbeidsgeschiktheid is georganiseerd in vijftien landen en hoe de kwaliteitsborging is geregeld. Dit is gedaan met literatuuronderzoek, interviews ter plaatse en een vragenlijst. Het blijkt dat wettelijke criteria tussen de landen veel gemeen hebben maar dat de organisatie van de uitvoering sterk verschilt. De borging van de kwaliteit van de beoordelingen gebeurt in het algemeen indirect en impliciet. In hoofdstuk 3 wordt van vier landen beschreven hoe het redeneren van verzekeringsartsen er uitziet uitgaande van een casus van een oudere bouwvakker die arbeidsongeschiktheid claimt. Dat onderzoek is uitgevoerd met focusgroepen, gebruik makend van een schriftelijke casus. Daarna is een vragenlijst uitgezet bij een grotere groep verzekeringsartsen

in de vier deelnemende landen. De gronden voor de beoordeling, dat zijn de achterliggende aspecten die beoordeeld worden, die door verzekeringsartsen gebruikt worden blijken te passen bij de gebrekkigenrol en bij een eerlijk proces. Over het bepalen van de functionele mogelijkheden was grote overeenstemming, over de vraag wat van iemand gevergd kon worden alvorens blijvende ongeschiktheid aan te nemen was dat veel minder het geval. In hoofdstuk 4 wordt beschreven welke definitie van arbeidsongeschiktheid wordt gehanteerd door verzekeringsartsen in dertien landen en welke richtlijnen daarvoor eventueel bestaan. Dat is onderzocht met een vragenlijst die is voorgelegd aan centrale medisch adviseurs van de sociale verzekering in die landen en met interviews in de landen waarvan was gesteld dat er richtlijnen werden gebruikt. Vier landen blijken officiële richtlijnen te gebruiken, welke deels medisch inhoudelijk van aard zijn en deels procedureel. Medisch inhoudelijke richtlijnen geven invulling aan het beoordelen van aspecten van de gebrekkigenrol, procedurele richtlijnen aan een eerlijk proces. In hoofdstuk 5 wordt beschreven in welke mate de medisch inhoudelijke richtlijnen van Duitsland en Nederland op een verantwoorde manier ontwikkeld zijn. Dat wordt gedaan aan de hand van een voor de klinische praktijk ontwikkeld instrument: het AGREE instrument. De getoetste richtlijnen blijken in een aantal opzichten de toets goed te doorstaan maar niet in het betrekken van relevante partijen bij de ontwikkeling, in strakheid van de wetenschappelijke onderbouwing en in de onafhankelijkheid bij de ontwikkeling van de richtlijnen. In hoofdstuk 6 wordt beschreven welke protocollen in Nederland zijn ontwikkeld voor het voeren van de beoordelingsgesprekken. Dat wordt gedaan aan de hand van de literatuur over de protocollen zelf, in interviews met de opstellers van de protocollen en met de literatuur over de in de protocollen gehanteerde uitgangspunten. Er blijken in Nederland drie protocollen te zijn die vooral op praktijkervaringen zijn gebaseerd. Deze protocollen worden vergeleken op twee punten: 1) wat ze aanbevelen om in het gesprek aan de orde te stellen en 2) de manier waarop het gesprek volgens protocol moet worden vorm gegeven. Er werden veel overeenkomsten gevonden in wat er aan de orde moet komen en in de wijze waarop. Daarnaast werden veel verschillen in de meer gedetaillerde uitwerking gevonden, met name in de strakheid van de protocollen en bij de vraag of wel of niet activering tot werkhervatting of uitbreiding ervan moet worden nagestreefd of alleen beoordeling van mogelijkheden en beperkingen. In hoofdstuk 7 wordt beschreven in welke mate Nederlandse verzekeringsartsen de protocollen zeggen toe te passen waarin ze getraind zijn. Dit is gedaan met een vragenlijst onder honderdvijftig verzekeringsartsen. Het blijkt dat alle verzekeringsartsen een vorm van protocol gebruiken, hetzij een gepubliceerd dan wel een zelf opgesteld protocol. Er is geen significant verband tussen het getraind zijn in een bepaald protocol en het toepassen ervan. Er was overeenstemming bij meer dan 80% van de respondenten over de thema's in het gesprek die aan de orde moeten komen, die betrekking hebben op de gebrekkigenrol teneinde deze goed te kunnen beschrijven en beoordelen.

In hoofdstuk 8 worden de resultaten samengevat en van methodische kanttekeningen voorzien. Zo is de praktijk van de beoordeling niet in die praktijk zelf onderzocht. Verder is alleen gekeken naar eerste beoordelingen van arbeidsongeschiktheid, niet naar herbeoordelingen en niet naar bijzonder complexe gevallen. Wat betreft de procedurele richtlijnen (en in mindere mate medisch inhoudelijke richtlijnen) is de opbrengst van het onderzoek mogelijk een onderschatting van de werkelijke situatie. Er is een vermoeden dat procedurele richtlijnen in de praktijk meer worden gebruikt dan uit dit onderzoek blijkt.
Geconcludeerd wordt dat de beoordelingen van arbeidsongeschiktheid beoordelingen zijn van het naleven van de gebrekkigenrol door de cliënt. Dat doet recht aan zowel wettelijke als beleidsdoelstellingen en medische inzichten, onder andere van de verzekeringsartsen zelf. Verder blijkt het nastreven van een eerlijk proces een belangrijk criterium voor het vormgeven van de beoordelingen. Dat past bij het uitvoeren van een wet. De definitie van kwaliteit en de operationalisatie ervan is ook bij de beoordeling van arbeidsongeschiktheid primair aan de experts. Die hebben daar in een aantal landen het een en ander aan gedaan, met name door het opstellen van professionele richtlijnen. Duidelijk is ook dat de definitie van kwaliteit en de borging ervan niet alleen aan de experts is maar ook aan hun opdrachtgever en aan de cliënten als geheel. Daar is nog veel winst te behalen.
Aanbevolen wordt dan ook om de experts enerzijds meer uit te dagen en anderzijds meer te faciliteren om hun verantwoordelijkheid in deze te nemen. Het maken van en verbeteren van medische en procedurele richtlijnen is daar een voorbeeld van, evenals het verder ontwikkelen van de principes van een goede gespreksvoering en mediprudentie. Dat alles behoeft een stevige wetenschappelijke onderbouwing, want voorlopig geldt dat in de verzekeringsgeneeskunde aanzienlijk meer is overdacht dan dat er is onderzocht.

Dankwoord

Drie keer is scheepsrecht en dit proefschrift is dan ook mijn derde poging daartoe. Over die pogingen zijn 22 jaar heen gegaan. Dat het allemaal zo lang heeft geduurd heeft alles te maken met het feit dat ik zoveel dingen de moeite waard vind en zo slecht nee kan zeggen tegen een leuk initiatief. Zelfs als ik het zelf heb verzonnen.

Maar deze poging lukt en ik wil het proefschrift niet afsluiten zonder een aandacht te besteden aan mensen die cruciaal waren bij deze derde poging. Dat zijn er meteen teveel om op te noemen. Honderden respondenten hebben de diverse onderzoeken mogelijk gemaakt door belangeloos mee te werken. Tientallen collega's zijn betrokken geweest hetzij bij de opzet en uitvoering van de deelonderzoeken en het schrijven van de artikelen dan wel als collega's bij TNO en KCVG en elders die mijn kennis en motivatie op peil hielden. Al deze mensen dank ik hartelijk.

Aan de wieg van dit proefschrift staat wijlen Hans van Oijen die heeft bevorderd dat ik er aan zou beginnen en het SIG bestuur heeft overtuigd van de wenselijkheid om daar ook geld in te steken. Ik betreur het bijzonder dat Hans het niet meer kan meemaken. Bij TNO zijn mijn opeenvolgende bazen Kees Wevers, Dick van Putten en Jan Besseling bereid geweest budget voor het onderzoek vrij te maken. Jan heeft me op zijn gemoedelijke manier al die tijd aangespoord er prioriteit aan te geven en hij heeft er soms meer in geloofd dan ik zelf. De organisaties SIG en TNO ben ik zeer erkentelijk voor de verleende financiële steun.

Iemand van wie ik het ook zeer betreur dat hij het niet meer meemaakt is Henk Herngreen. Zijn ideeën omtrent claimbeoordeling, argumentatieve beoordeling en het forensische karakter van de verzekeringsgeneeskunde hebben mijn denken sterk gestuurd. Voor een belangrijk deel komen ze in dit proefschrift terug. Henk zou zeker nog heel veel aan te merken hebben op het product maar hij zou er toch vast ook blij mee zijn geweest. Aan hem draag ik dit proefschrift op, in dankbare nagedachtenis zoals dat dan heet.

Veel van de activiteiten in het kader van dit proefschrift vergden medewerking van leden van mijn beroepsgroep. Dat ik daarvan gebruik heb kunnen maken is mede te danken aan het vertrouwen dat het bestuur van de NVVG en het management van UWV in mij hebben gesteld. Daar dank ik ze voor.

Twee promotoren en een copromotor hebben gemaakt dat al dit onderzoek ook echt een proefschrift werd.
Han Willems was vanaf het begin betrokken bij de inhoud en de organisatie. Met zijn droge stijl van feed back geven heeft hij me constant uitgedaagd

er iets goeds van te maken maar ook mijn motivatie goed onderhouden. Als iemand de hoofdlijn bewaakt heeft is Han het geweest. Alsmaar de berg op en weten waar je naartoe wilt en dan zo simpel mogelijk. Dank daarvoor, je weet hoe makkelijk ik verdwaal in mijn eigen associaties. Dat heb je geen moment getolereerd.
Peter Donceel heeft als EUMASS- collega, als co- auteur en later formeel als promotor de internationale kanten bewaakt, ook bestuurlijk. Het was en is altijd een feest om over de grenzen te kijken en met buitenlandse collega's de gemeenschappelijke problemen te bespreken en te onderzoeken. Peter heeft altijd voorwaarden daarvoor geschapen; hopelijk blijven we dat doen.
Frank van Dijk kende ik als een echte bedrijfsarts- onderzoeker. Nu ken ik hem als meer. De methoden en discussies van de verzekeringsartsen heeft hij zich met zijn gebruikelijke enthousiasme eigen gemaakt en bestookt met kritisch commentaar. Frank legde een soms wanhopig makend vermogen aan de dag om bij de volgende versie vergeten te zijn wat hij eerder had gezegd; iedere versie werd geheel als nieuw beoordeeld wat bijzonder verfrissend is, ook.

De promotiecommissie is van een voor mij zeer eervolle samenstelling. Graag dank ik hen voor de moeite die ze genomen hebben om dit werk tot zich te nemen en de vriendelijke commentaren die ik al mocht ontvangen. Ik zie uit naar de vragen en hoop met mijn antwoorden niet teleur te stellen.

Ik ben wel een familiemens. Mensen zijn kuddedieren en dat zie ik niet voor niets als biologische basis voor de sociale verzekering en daarmee voor de claimbeoordeling. Ik ben dan ook blij dat een (toegegeven: heel kleine) selectie uit de familie, mijn broers Dick en Rob, de ceremonie luister wil bijzetten als paranimfen.

Scheiding van werk en privé is in dit project niet erg gelukt. Frank, Anke en Eva hebben met hun aanwezigheid en hun vragen en af en toe spottende commentaren me constant herinnerd aan waarom ik het ook weer allemaal doe. Anke heeft ook meegeholpen met duwen aan het werk zelf. Zowel hun bestaan als hun gedrag vervullen me met dankbaarheid. Jacqueline heeft alle pogingen en ups en downs, de kloven!, van zeer nabij mee gekregen. Je steun was onwankelbaar en dat heb ik veel vaker nodig gehad dan me lief was. Tijdens deze derde poging heb je me gelukkig ook op een ander been gezet en in de Argentijnse Tango ontwikkelen we ons samen steeds verder. Zwierend naar de einder…

Scientific Publications

Albertsma T. & Boer W.E.L. de. **Argumentatie in de verzekeringsgeneeskunde**. *TVG* 1991 **29**: 142-144

Foekens R. & Boer W.E.L. de. **Doelen van oordeelsvorming in de verzekeringsgeneeskunde**. *NTvG* 1991 *135*: 1892-1896

Stoep J. van der & Boer W.E.L. de. **Verhoudingsgewijs gering gebruik door allochtone inwoners van verplaatsingsvoorzieningen voor gehandicapten**. *NTvG* 1992 **136**: 1062-1065

Boer W.E.L. de. **Argumentatieve claimbeoordeling I: de claim**. *TVG* 1992 **30**: 110-114

Boer W.E.L. de. **Argumentatieve claimbeoordeling II: de argumentatie**. *TVG* 1992 **30**: 137-139

Boer W.E.L. de. **Argumentatieve claimbeoordeling III: de beoordeling in gesprek**. *TVG* 1992 **30**: 171-176

Boumen W.M.M.L., Boer W.E.L. de. **Zwangerschap en Ziektewet: een onderzoek naar de gevolgen van gewijzigde regelgeving**. *TBV* 1993 **1**: 92-94

Boer W.E.L. de. Giezen A.M. van der, Josée L.H., Wieten H., Til M.J. van. **Gegevensoverdracht van arbodienst naar bedrijfsvereniging**. *TBV* 1995 **3**: 46-50

Boer W.E.L. de, Croon N.H.Th., Toorn M. van den, Csanky-Achilles H.W.. **Enquête medisch arbeidsongeschiktheidscriterium**. *TBV* 1996 **4**: 98-103

Boer W.E.L. de, Wijers J.H.L., Spanjer J., Beijl I. van der, Zuidam W., Venema A. **Gespreksmodellen in de verzekeringsgeneeskunde**. *TBV* 2006 **14**: 17-23

Kerstholt J.H., Boer W.E.L. de, Jansen N.J.M. **Disability assesments: effects of response mode and experience**. *Dis & Rehab*. 2006 **28**: 111-115 2006

Boer W.E.L. de, Besseling J.J.M, Willems J.H.B.M., **Organisation of disability evaluation in 15 countries**. *Pratiques et Organisation des Soins* **38**: 205-217 2007

Boer W.E.L. de; Brage S. **Dilemmaer i legers uførhetsvurderinger - et europeisk perspektiv** [Dilemmas in medical evaluation of work disability; a European perspective] *Tydsskrift for velferdsforskning* **11**: 104-112 2008

Latil F., Boer W.E.L. de, Cardona J. **Impact of the no tort reform on medical liability** *[in French] POS* **39**: 331-339 2008

Boer W.E.L. de, Donceel P., Brage S., Rus M., Willems J.H.B.M. **Medicolegal reasoning in disability assessment: a focus group and validation study**. *BMC Public Health 2008* **8**: 335

Boer W.E.L. de, Wind H., Dijk F.J.H. van, Willems J.H.B.M. **Interviews for the assessment of long-term incapacity for work: a study on adherence to protocols and principles**. *BMC Public Health* 2009 **9**: 169

Boer W.E.L. de, Rijkenberg A.M,, Donceel P. **Guidelines in assessment of work disability: an international survey**. Submitted *Das Gesundheitswesen*

Boer W.E.L. de, Bruinvels D.J., Rijkenberg A.M., Donceel P., Anema J.R. **Evidence based guidelines in the assessment of disability for work: quality of development according to AGREE**. *BMC Public Health* **9**: 349 2009

Muizer A.M., Brouwer S., Boer W.E.L. de, Geertzen J.H.B., Groothoff J.W. **The assessment of efforts to return to work: a Dutch – European comparison**. Submitted *EJPH*