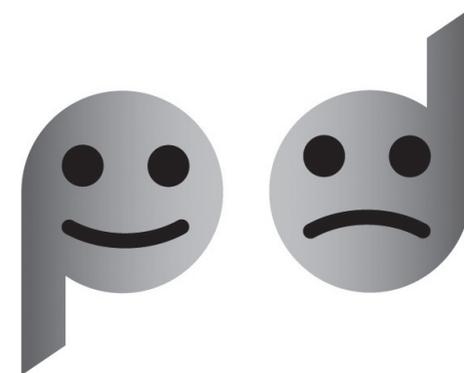


# **Development and evaluation of an implementation strategy for insurance medicine guidelines for depression**



VRIJE UNIVERSITEIT

# Development and evaluation of an implementation strategy for insurance medicine guidelines for depression

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van Doctor  
aan de Vrije Universiteit Amsterdam,  
op gezag van de rector magnificus  
prof.dr. L.M. Bouter,  
in het openbaar te verdedigen  
ten overstaan van de promotiecommissie  
van de Faculteit der Geneeskunde  
op dinsdag 29 januari 2013 om 13.45 uur  
in de aula van de universiteit,  
De Boelelaan 1105

door

**Feico Zwerver**

geboren te Alkmaar

The study presented in this thesis was conducted at the EMGO+ Institute for Health and Care Research, Department of Public and Occupational Health of the VU University Medical Center Amsterdam, The Netherlands. The Department of Public and Occupational Health of the VU Medical Center participates in the Dutch Research Center for Insurance Medicine, which is a joint initiative of the VU University Medical Center (Department of Public and Occupational Health, EMGO+ Institute for Health and Care Research), Amsterdam Medical Center, the University Medical Center Groningen, and the Dutch Institute for Employee Benefits Schemes (UWV). The EMGO+ Institute participates in the Netherlands School of Primary Care Research (CaRe), which was re-acknowledged in 2005 by the Royal Dutch Academy of Arts and Sciences (KNAW). The experiment in this thesis was carried out in collaboration with the Netherlands School of Public & Occupational Health (NSPOH).

This study was funded by the Dutch Institute for Employee Benefits Schemes (UWV).

Financial support for the printing of this thesis has been kindly provided by the UWV: 'divisie HRM, afdeling opleidingen'.

**English title:** Development and evaluation of an implementation strategy for insurance medicine guidelines for depression

**Nederlandse titel:** Ontwikkeling en evaluatie van een implementatie strategie voor het verzekeringsgeneeskundig protocol Depressieve stoornis

## Copyright

© 2012, Feico Zwerver

All rights reserved. No parts of this book can be reproduced in any form without written permission from the author.

## ISBN/EAN

978-90-6464-611-9

## Cover design

Jacques Koeweiden, Amsterdam ([www.koeweidenpostma.nl](http://www.koeweidenpostma.nl))

Cover photo: actress Ilke Turpijn; stills taken from one of the DVD's used in this research project.

## Layout and printing

GVO drukkers & vormgevers B.V. | Ponsen & Looijen

promotoren: prof.dr. A.J. van der Beek  
prof.dr. J.R. Anema

copromotor: dr. A.J.M. Schellart

## Contents

<b>Chapter 1</b>	General introduction	<b>7</b>
<b>Chapter 2</b>	Intervention mapping for the development of a strategy to implement the insurance medicine guidelines for depression <i>BMC Public Health</i> 2011, <b>11</b> :9	<b>19</b>
<b>Chapter 3</b>	Development and reliability of performance indicators for measuring adherence to a guideline for depression by insurance physicians <i>Disabil Rehabil</i> 2011, <b>33</b> :2535-43	<b>43</b>
<b>Chapter 4</b>	Relationships between intention to use guidelines and determinants of insurance physician's behaviour <i>Submitted</i>	<b>61</b>
<b>Chapter 5</b>	An implementation strategy for insurance physicians to improve guideline adherence: an experiment in a controlled setting <i>Implement Sci</i> 2011, <b>6</b> :131	<b>79</b>
<b>Chapter 6</b>	Changes in insurance physicians' attitudes, self-efficacy, intention, and knowledge and skills regarding the guidelines for depression, following an implementation strategy <i>J Occup Rehabil</i> 2012, 5 jul.	<b>99</b>
<b>Chapter 7</b>	The influence of applying insurance medicine guidelines for depression on disability assessments <i>Submitted</i>	<b>119</b>
<b>Chapter 8</b>	Implementation of a strategy for the insurance medicine guidelines for depression: a process evaluation <i>Submitted</i>	<b>135</b>
<b>Chapter 9</b>	General discussion and conclusions	<b>153</b>
<b>Summary</b>		<b>173</b>
<b>Samenvatting</b>		<b>181</b>
<b>Dankwoord</b>		<b>189</b>
<b>About the author</b>		<b>197</b>

# **Chapter**

---

**1**

## **General Introduction**

In the year 1990 the Prime Minister announced that “The Netherlands are ill” as the number of employees receiving disability benefits reached 900,000, testing the limits of the nation’s social security. Since then, disability benefits drew serious political attention, followed by series of adjustments in work disability legislation. Furthermore, the focus was on the performance of the professionals working in the field of insurance medicine. The question of who was to blame for the immense increase of disabled employees arose. Was it the employees themselves, the employers, economics, or the insurance physicians working at the Dutch Institute for Employee Benefits Schemes (Institute). Were these 900,000 employees really ill? To what standards did insurance physicians (IPs) actually work?

### **Work disability assessment by an insurance physician in the Netherlands**

Employees who are on sick leave for two years can claim a disability benefit through the Institute. Such an employee becomes a client of the Institute. The clients’ claim is assessed by an IP at a front office of the Institute. In this assessment, that is called the work disability assessment, the client’s work limitations and abilities are defined. The IP writes his or her findings down in a medical work disability report and fills in a Functional Ability List (FAL). On average, an IP uses approximately two hours for a complete work disability assessment. One hour for the assessment interview, and one hour for writing the report. Subsequently, a labour expert matches the client’s work abilities with the functional demands of (theoretically) available jobs, resulting in a selection of jobs that the client should be able to perform, despite his/her work limitations. The client’s benefit, finally, is determined by the loss of income, caused by the difference in wages between that of the client’s initial job and the wages of the selected jobs. During the two years of sick leave previous to the visit to the IP, the client has been attended to by an occupational physician. At the end of the two-year period of sick leave, the occupational physician transfers the client to the Institute, and supplies the IP with medical information. Although the work disability assessments of clients traditionally are being executed under social security legislation, specific guidelines for the IPs were lacking until the year 2007.

### **Guideline implementation in insurance medicine in the Netherlands**

In 2006 the Minister of Social Affairs directed the Health Council to develop insurance medicine guidelines. In response, within two years, between 2007 and 2009, 20 different



insurance medicine guidelines were developed by the Health Council and the Dutch Association of Insurance Medicine. These guidelines covered diseases contributing most to the total number of disabled employees. The 20 guidelines were distributed and implemented at the Institute within this two-year period. Although these guidelines were supposed to be evidence-based, their implementation at the Institute was not evidence-based. It was not unusual then, that two different guidelines were implemented in one afternoon session. A needs assessment that was carried out among IPs, staff IPs and stakeholders of the Institute showed that there was no experience with implementing guidelines. Staff IPs and stakeholders did not know what was the best way to do it. IPs on their side felt difficulties with using guidelines in practice. They were not used to working with guidelines in practice. Most of them had been practising for years without using guidelines. There was neither monitoring of IPs’ performances in the use of the guidelines nor any evaluation of the use of guidelines in practice. IPs wanted to know whether they could be facilitated in applying the guidelines. This needs assessment made clear that implementing guidelines appeared to be a problem at the Institute. The Institute had to implement the guidelines, was left with no choice, except for starting research to the implementation of guidelines. Questions arose: how can it be done at best, and will the IPs then adhere to guidelines? Research to the implementation of guidelines in the field of insurance medicine was definitely needed.

Resuming, there was a need from the IPs as well as from the stakeholders for an effective implementation strategy of insurance medicine guidelines. In this thesis we tried to meet the needs of the stakeholders as well as the needs of the IPs, in the development of an implementation strategy concerning the guidelines for depression. This implementation strategy aimed to make it easier for the IPs to use the guidelines in practice.

### **Work disability assessment by an insurance physician according to the guidelines**

The 20 developed insurance medicine guidelines have a general introduction to these guidelines in common [1]. In this general introduction to the guidelines, the framework of the guidelines and the contemplated application are explained.

In general, when assessing a client the IP has to carry out four main tasks:

1) The assessment of the social medical case history of the client.

The essence of this assessment is the analysis of client’s stagnation in recovery and return to work.

2) The assessment of the actual functional abilities of the client.

This is the main task of an IP. The IP has to assess to which degree the clients' functional abilities are limited by disease, disorder or defect.

3) The assessment of the expected course (prognosis) of the disease and the prognosis of functional abilities of the client.

The prognosis of the disease or disorder has importance for the assessment of the durability of total work disability, the assessment of treatment and therapy and for the planning of re-assessments.

4) The evaluation of the current and initiated treatment and therapy.

According to the guidelines, an IP has to analyse and assess the return to work activities that have been carried out, either by the client, the employer, or professionals such as the occupational physician from the occupational health service, who has been attending to the client in the first two years of sick leave. Furthermore, the IP has to assess treatment and therapy the client has received from his or her curative physicians. In particular, the reasons for not recovering and not returning to work have to be assessed.

### **Work disability assessment by an insurance physician according to the guidelines for depression**

Among the 20 guidelines the guidelines for depression [2] were chosen to be the subject of this thesis, because of its societal relevance. Worldwide depression is a major health problem. According to the WHO, in 2020 depression is expected to take the second place as a cause for lost disability-adjusted life-years, after cardiovascular disease [3-5]. The individual with a depression is not only suffering from depressive symptoms, such as a depressed mood or loss of interest or pleasure, but might also show hampered social functioning at home or limited performance at work [6-8]. Besides, depression is regarded as a chronic disease, tending to relapse [2]. Therefore, depression is a disorder, which is highly important for insurance medicine. Assessing work disability of mental illness in general is found to be difficult for IPs, probably because mental illnesses are not as open to objectification as somatic diseases are [9]. Guidelines might meet the IPs' uncertainties in managing mental illness by supporting them with evidence-based medicine [10].

The insurance medicine guidelines for depression contain six main elements:

1) The assessment of the diagnosis according to the Diagnostic Statistic Manual for Mental Disorders (DSM-IV).

In this manual the symptoms of depression are listed. A client should have at least five out of the nine listed symptoms, before the diagnosis depression can be made.



2) The assessment of the severity of the disorder.

In case of depression the severity of the disorder and the work limitations are positively associated. Having more severe symptoms of depression influences the work limitations.

3) The assessment of the course of the disorder, and the presence of risk factors for the disorder.

The course of the disorder is important, because depression often acts as a chronic disease. Having insight in the presence of potential risk factors for depression is needed for predicting the course and assessing the prognosis of the disorder for a certain client.

4) The assessment of co-morbidity.

The presence of existing co-morbidity also determines the work abilities of the client. Sixty percent of the adult people with depression has psychiatric co-morbidity [2].

5) The evaluation of treatment and therapy.

Clients do not always receive evidence-based treatment. Or their compliance with evidence-based therapy is not always optimal. The IP analyses the reasons for stagnation in their recovery and return to work.

6) The assessment of the work limitations caused by the disorder.

Clients with depression can have specific work limitations, such as having problems with their attention, or having difficulties with executing complex tasks, or having problems in social interaction with other people at work.

A short case history illustrates the difficulties insurance IPs might face in the assessment of a client suffering from depression.

### **Case**

*A 53-year old teacher at a secondary school is on sick leave for almost two years.*

*He has complaints of fatigue, is having problems with his concentration, and does not feel like doing anything anymore. The IP investigates about his work situation: he has a discipline problem with the students, a communication problem with some of his colleagues, he lacks back-up by the staff after dismissing a student, and he can not support nowadays teaching methods anymore. He says he is a good traditional teacher, but those students do not listen to him anymore. He feels guilty, not working, and at the same time he does not like the idea of going back to work at all. He does not know what to do.*

*Private situation: his wife suffers from a manic depression and, in the past, he once found her bleeding in the bath. His mother suffers from Alzheimer and does not recognize him any more. He visits her every week. He tends to isolate himself. "If I wouldn't wake up in the morning, I wouldn't care", he sighed.*

*Symptoms: feeling deeply sad, 10 kg of weight loss since sick leave, sleeping problems.*

*Personality characteristics: some features of obsessive compulsive disorder.*

Suppose that this teacher's work ability was assessed by a random IP of the Institute.

The conclusion of this IP might be: this teacher just shows a normal reaction on having hard times. This man is not ill. He should go back to school and return to work. No benefits.

If this case has been given a second opinion, the conclusion of this IP or another random IP at the Institute could also be: this teacher suffers from a major depression. He is, therefore, not able to work anymore and considering the risk factors at work and at home he has a bad prognosis for recovering. Benefits agreed.

This example of the teacher, although exaggerated, makes clear that it might be difficult for IPs to assess work disability of clients with mental illness, and that specific guidelines are needed. As is shown in this example a lack of uniformity between IPs' policies in work disability assessments might occur easily. Although each conclusion could be justified, the difference between the outcomes of this assessment is not desirable from the perspective of 1) the client, 2) the IP, 3) the stakeholders of the Institute, and 4) Dutch society.

Ad 1) The client deserves careful evidence-based assessment according to the guidelines, and not to be at someone's mercy. If an IP treats the client according to proper guidelines, then the assessment will be more evidence-based, which means that all aspects of the sick leave period, ranging from client's history before becoming ill and assessment of the diagnosis to the revalidation and participation, are paid attention to. For instance, has everything been done to have the teacher stay at work? Then the client might feel being taken seriously.

Ad 2) IPs will benefit from evidence-based assessments in accordance with guidelines; their work will be more transparent. Transferring a client from one IP to another will be easier when the IPs' policy concerning a certain client is transparent, evidence-based and laid down in a well-argued written report. This is particularly important in appeal cases. Furthermore, one of the IPs' roles is that of 'gatekeeper' to the inflow into work disability benefits. Not every sad feeling client suffers from depression according to the DSM-IV. The guidelines for depression support IPs in separating "the sheep from the goat" as it comes to assessing the diagnosis depression. Unwanted low agreements



(intra-IP and inter-IP) between the work disability assessments of a client might be positively influenced by guidelines.

Ad 3) Stakeholders at the Institute and policymakers have interest in uniformity, transparency and quality in the execution of work disability assessments, since they are responsible for the performance of social security in commission of the Ministry of Social Affairs. Too much IP disagreement or arbitrariness in the award of disability benefits might cause negative reactions in public opinion. Besides, several patient associations critically follow the impact of insurance medicine guidelines, standing up for the sake of their clients.

Ad 4) Dutch society expects careful and fair work disability assessments of employees claiming benefits.

## Guideline adherence

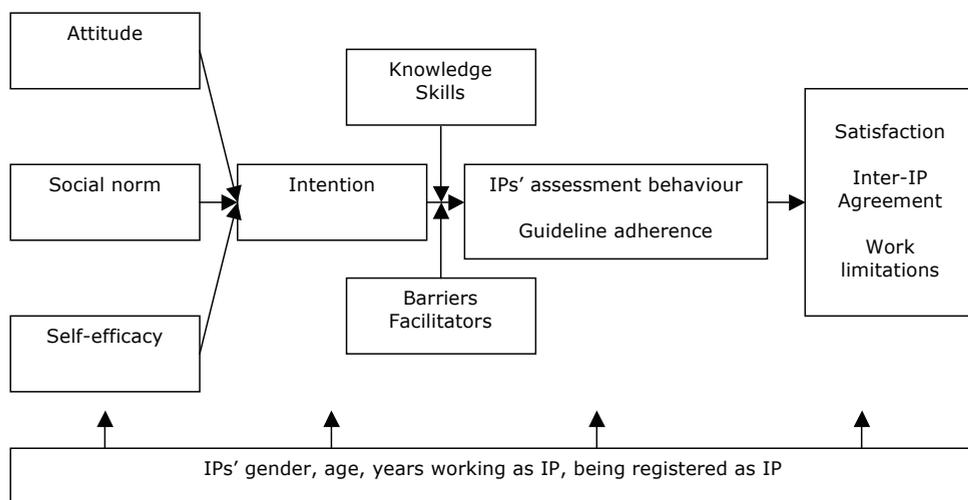
As earlier illustrated clients, physicians, stakeholders, policymakers and society could benefit from guidelines. Guidelines are considered to be one of the major efforts to improve quality of care, implying that physicians should use guidelines in daily practice. However, another problem arises; from research it is known that physicians not always completely adopt guidelines in their working routines. On average patients receive 55% of the recommended care ranging from 78% for senile cataract to 10% for alcohol dependency [11]. Reasons for not following guidelines by physicians were investigated in several studies [12-16]. Barriers to successful implementation of guidelines were described at different levels, such as patient's level, organizational level, or physician's level [17, 18]. It becomes obvious that implementing guidelines is a difficult and challenging task. In the Netherlands, occupational physicians preceded IPs in the implementation of guidelines by approximately five years, while general practitioners celebrated the twentieth anniversary of their guidelines with a national symposium in 2009 [19, 20].

## Conceptual model

Evaluating the implementation of guidelines implies research into the concept of physicians' guideline adherence. Physicians' guideline adherence can be regarded as behaviour of professionals concerning the use of guidelines. The Theory of Planned Behaviour (TPB) is designed to predict and explain human behaviour in specific contexts [21]. In this thesis the Attitude, Social-influence, self-Efficacy (ASE) model, a derivative

from the TPB, was used for describing the behavioural determinants of physicians' guideline adherence [22]. Motivational factors to perform a given behaviour are captured by intentions to perform that behaviour. Intentions, in turn are preceded by attitude, social norm and self-efficacy concerning the desired behaviour. Applying the ASE model to guideline adherence of physicians, its concepts have the following meaning: guideline adherence is the IPs' behaviour towards the use of guidelines. IPs are thought to have a certain attitude (positive or negative) towards guidelines, which influences their intention to use them. Social influence is the opinion of the IPs' colleagues or staff or their Medical board concerning the use of guidelines. The IPs' intention to use guidelines could be influenced by their colleagues or staff. If colleagues do not use guidelines and staff does not stimulate or evaluate the IPs' performance on using them, why should an individual IP use guidelines? The concept of self-efficacy is the perception of behavioural control, felt by an IP in case of applying the guidelines. Hence, self-efficacy has an influence on the IPs' intentions too. The relationships between the determinants of behaviour, such as attitude, social-influence, self-efficacy, intention, and the interfering facilitators or barriers to perform the expected behaviour are shown in the ASE model, illustrated in Figure 1.1. For the purpose of this thesis, the ASE model was extended with the outcomes of this research project, which all are related to the behaviour of the IPs: inter-IP agreement in the Functional Ability Lists (FAL), and satisfaction with working according to the guidelines.

More information on the IPs' behaviour described by the ASE model can be found in Chapter 4 and 6 of this thesis.



**Figure 1.1:** ASE model adjusted to this thesis. Model describing insurance physicians' assessment behaviour in case of using the guidelines for depression and the outcomes. IP=Insurance Physician.



## Questions

Question asked by stakeholders of the Institute and IPs:

### **Which strategy can be developed to implement the guidelines for depression, in order to promote use by IPs?**

For IPs and the stakeholders of the Institute it is important to close the gap between evidence-based medicine and daily practice. Stakeholders and IPs wanted to know whether the developed implementation strategy is feasible in every day practice, and whether it will contribute to quality of care. Guidelines should be implemented preferably without loss of production. Furthermore, stakeholders and policymakers have an interest in the outcomes of the work disability assessments by the IPs. In general, do changes in guideline adherence of IPs lead to changes in the outcomes of the work disability assessments? Will improvements in guideline adherence support uniformity, and will the inter-IP agreement in assessed work abilities between individual IPs improve as well? Given a certain client, if an IP follows guidelines more strictly, what kind of influence will there be on the work abilities for that client? In other words, will the total volume of work disability change due to the implementation of guidelines? IPs share afore-mentioned interests of the stakeholders and policymakers, because IPs have a broad societal common sense, and therefore need feedback on the outcomes of their work. We translated these questions into:

**What are the effects of such a developed implementation strategy on the guideline adherence of the IPs and on their knowledge of the guidelines?**

**What are the effects of such a developed implementation strategy on the behavioural determinants of the IPs regarding the use of the guidelines?**

**What are the effects of such a developed implementation strategy on the number and severity of work limitations when applying the guidelines?**

**What are the effects of such a developed implementation strategy on the inter-IP agreement in the work disability assessments of the IPs?**

**What are the effects of such a developed implementation strategy on the satisfaction of the IPs?**

We thought if we could manage to develop an implementation strategy for the guidelines for depression that suits to the IPs' practice, the next step is to scientifically evaluate

this strategy on the outcomes: IPs' behaviour towards guidelines, guideline adherence, satisfaction with the developed strategy, the inter-IP agreement in the work disability assessments, and work limitations.

## Objectives and outline of the thesis

The main objective of this thesis is:

### To develop and evaluate a multifaceted implementation strategy for the insurance medicine guidelines for depression

Chapter 2 describes how we developed the implementation strategy by taking into account the IPs' and stakeholders' needs by the use of Intervention Mapping. Intervention Mapping facilitates the development of the implementation strategy, by integrating the needs of the IPs and the stakeholders with evidence-based medicine and expert opinions in a framework. For each development of tools, performance indicators and the training, we used separate expert groups. The results from the expert groups are briefly presented in this chapter.

In Chapter 3, the performance indicators are outlined in more detail, from the development, to content validity and reliability. The behaviour of IPs regarding the guidelines is explored and described in Chapter 4. Changes in that behaviour due to the implementation strategy are reported in Chapter 6. Furthermore, in Chapter 6 the changes in IPs' behaviour were linked to changes in observed guideline adherence, which are presented in Chapter 5.

Chapter 5 describes the effects of the implementation strategy on the guideline adherence of the IPs and their knowledge of the guidelines by presenting the results of a controlled experiment. The effects of the implementation strategy on the guideline adherence in the IPs' work disability reports were measured using performance indicators.

We reported on the influence of the implementation strategy on the number and severity of work limitations and the inter IP-agreement in the disability assessments in Chapter 7. The experiences, satisfaction and expectations of the IPs with the developed implementation strategy are shown in the process evaluation, Chapter 8. Finally, Chapter 9 presents the general discussion of this thesis. In this chapter answers will be given to the questions asked in the general introduction, which content the main findings of this thesis. Furthermore, methodological and theoretical considerations, practical implications, and recommendations for future research are discussed.



## References

1. Health Council: *General introduction to the guidelines for insurance medicine*. Den Haag: Gezondheidsraad, 2006.
2. Health Council: *Insurance medicine guidelines for depression*. Den Haag: Gezondheidsraad, 2006.
3. Bijl RV, Ravelli A, Van Zessen G: Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Soc Psychiatry Psychiatr Epidemiol* 1998, **33**:587-95.
4. Murray CJ, Lopez AD: Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet* 1997, **349**:1498-1504.
5. Brown P: Effective treatments for mental illness are not being used, WHO says. *BMJ* 2001, **323**:769.
6. Lerner D, Henke RM: What does research tell us about depression, job performance, and work productivity? *J Occup Environ Med* 2008, **50**:401-10.
7. Goetzel RZ, Ozminkowski RJ, Sederer LI, Mark TL: The business case for quality mental health services: why employers should care about the mental health and well-being of their employees. *J Occup Environ Med* 2002, **44**:320-30.
8. Eaton WW, Martins SS, Nestadt G, Bienvenu OJ, Clarke D, Alexandre P: The burden of mental disorders. *Epidemiol Rev* 2008, **30**:1-14.
9. Linder J, Ekholm KS, Jansen GB, Lundh G, Ekholm J: Long-term sick leavers with difficulty in resuming work: comparisons between psychiatric-somatic comorbidity and monodiagnosis. *Int J Rehabil Res* 2009, **32**:20-35.
10. Williams CD, Schouten R: Assessment of occupational impairment and disability from depression. *J Occup Environ Med* 2008, **50**:441-50.
11. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, Kerr EA: The quality of health care delivered to adults in the United States. *N Engl J Med* 2003, **348**:2635-45.
12. Grol R: Successes and failures in the implementation of evidence-based guidelines for clinical practice. *Med Care* 2001, **39**:1146-1154.
13. Maue SK, Segal R, Kimberlin CL, Lipowski EE: Predicting physician guideline compliance: an assessment of motivators and perceived barriers. *Am J Manag Care* 2004, **10**:383-91.
14. Davis DA, Taylor-Vaisey A: Translating guidelines into practice. A systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines. *CMAJ* 1997, **157**:408-16.
15. Lomas J, Anderson GM, Domnick-Pierre K, Vayda E, Enkin MW, Hannah WJ: Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. *N Engl J Med* 1989, **321**:1306-11.
16. Grol R, Wensing M: What drives change? Barriers to and incentives for achieving evidence-based practice. *Med J Aust* 2004, **180**:S57-S60.
17. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al.: Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 1999, **282**:1458-65.
18. Hegarty K, Gunn J, Blashki G, Griffiths F, Dowell T, Kendrick T: How could depression guidelines be made more relevant and applicable to primary care? A quantitative and qualitative review of national guidelines. *Br J Gen Pract* 2009, **59**:e149-56.
19. Hulshof C, Hoenen J: Evidence-based practice guidelines in OHS: are they agree-able? *Ind Health* 2007, **45**:26-31.
20. Dutch College of General Practitioners, Utrecht: *NHG symposium 20 years of guidelines*. Amsterdam, 2009.
21. Ajzen I: The theory of planned behavior. *Organ Behavi and Hum Decis Process* 1991, **50**:179-211.
22. De Vries H, Dijkstra M, Kuhlman P: Self-efficacy: third factor besides attitude and subjective norm as a predictor of behavioral intention. *Health Educ Res* 1988, **3**:273-82.

# Chapter

---

# 2

## **Intervention mapping for the development of a strategy to implement the insurance medicine guidelines for depression**

BMC Public Health 2011, 11:9

**F Zwerver  
AJM Schellart  
JR Anema  
KC Rammeloo  
AJ van der Beek**

## Abstract

**Background:** This article describes the development of a strategy to implement the insurance medicine guidelines for depression. Use of the guidelines is intended to result in more transparent and uniform assessment of claimants with depressive symptoms.

**Methods:** The implementation strategy was developed using the Intervention Mapping (IM) method for alignment with insurance-medical practice. The ASE behavioural explanation model (Attitude, Social Influence and Self-Efficacy) was used as theoretical basis for the development work. A literature study of implementation strategies and interviews with insurance physicians were performed to develop instruments for use with the guideline. These instruments were designed to match the needs and the working circumstances of insurance physicians. Performance indicators to measure the quality of the assessment and the adherence to the guidelines were defined with input from insurance physicians.

**Results:** This study resulted in the development of a training course to teach insurance physicians how to apply the guidelines for depression, using the aforementioned instruments. The efficacy of this training course will be evaluated in a Randomized Controlled Trial.

**Conclusions:** The use of IM made it possible to develop guideline support instruments tailored to insurance medical practice.



## Background

Depression is an enormous health problem, which is responsible for 11% of disability worldwide [1]. The WHO predicts that by 2020, depression will be second only to heart disease as a cause of lost disability-adjusted life-years and untimely death [2]. Through social insurance, employees can claim compensation when they lose (part of) their income due to disability. To determine these disability benefit claims, disability assessments are carried out by specialized physicians, who have to evaluate the claimants' medical status and functional capacities with regard to vocational rehabilitation [3]. In the Netherlands, these assessments are performed by insurance physicians (IPs) who work for the Dutch Institute for Employee Benefit Schemes (Institute). The context of insurance medicine in the Netherlands is presented in Figure 2.1 [4]. Worldwide, physicians are involved in similar assessments, even though national practices, social systems and, disability legislation, may vary considerably [5].

The Dutch National Institute for Employee Benefits Schemes (Institute) administers the eligibility of sick employees for a benefit under the Work and Income (Capacity for Work) Act (WIA). 750 Insurance physicians are employed at the Institute, approx. 450 of who perform disability assessments under the WIA. On average, these insurance physicians are 50 years old, they are generalists, have approx. 16 years experience as insurance physician, approx. 85% is specialized in insurance medicine, 15% also has another extra medical speciality, and approx. 60% works full-time. They perform an average of 10 disability assessments per week, assessing patients with all types of diseases.

**Figure 2.1:** *Insurance physicians in the Netherlands; source: R. Steenbeek [4].*

In the Netherlands, 19 diagnosis specific guidelines, including depression, have recently been developed for use in insurance-medical practice [6, 7]. These guidelines are intended to serve as a reference framework that can help IPs to make their disability assessments more evidence-based and more standardized [8, 9]. IPs have to know all 19 guidelines and apply them in practice because they are generalists. The guidelines were subsequently implemented top-down by the Institute in the period between 2007 and 2009. This tight schedule of guideline implementation did not leave much time for the IPs to really apply all these guidelines. During this period, two guidelines were sometimes implemented in one single afternoon session. Unfortunately, little attention was paid to the needs of the IPs. Except for lack of time, the implementation of evidence-based guidelines in health care practice has proven to be difficult anyway [10]. Implementing guidelines requires the planning of complex changes in practice. Potential barriers at various levels need to be overcome, such as the nature of the guidelines, the characteristics of the physicians involved, and the social, organizational, economical and political context [11-14]. Using the intervention mapping method (IM),

it is possible to make provisions for the needs of the users and those around them, and to draw upon scientific theory and evidence, in the implementation of protocols and guidelines. Since 1998, IM has mainly been used for planning theory- and evidence-based health promotion programs [15-17]. However this method has now reached the field of occupational health medicine, where it is being used to support the development of intervention programs focussing on work disability [18-20]. This article describes the use of intervention mapping for the development of a strategy for the implementation of insurance medicine guidelines for depression. The aim was to answer the following core question: What approach should be taken to implement these guidelines, in order to ensure effective use by insurance physicians? A randomized controlled trial (RCT) will in due course be carried out to compare the efficacy of the implementation strategy described in this article with conventional implementation methods.

## Methods

IM, developed in the nineties by Bartholomew et al. [15, 21], is a planning instrument that maps out the development process of an intervention from the basic needs to the potential solution. IM provides a stepwise process for decisions, based on theory and evidence. It consists of six steps, which are presented in Figure 2.2. We used IM as basis for developing a strategy: to find a way in which to implement the guidelines for depression that suits the needs of the IPs.

### Step 1: Needs assessment

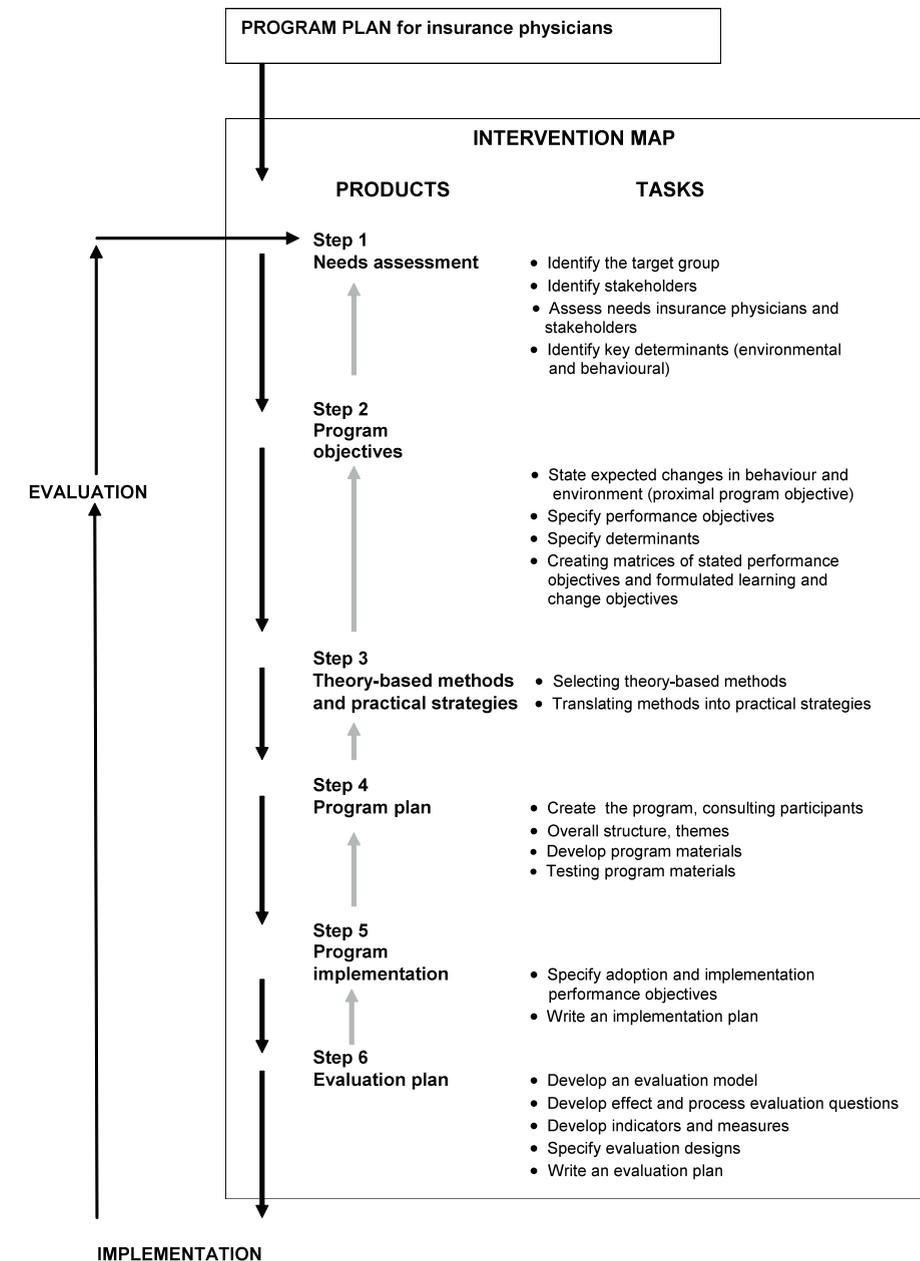
The key purpose of the needs assessment was to assess the needs of the IPs with regard to the guidelines for depression, as well as their opinions about the implementation of the guidelines at their place of work within the Institute. Interviews were held with 10 IPs working in practice (see Figure 2.3). They were asked to provide their opinions on: the content of the guidelines for depression, the possible obstacles, and the support, needed when using the guidelines in daily practice.

### Step 2: Program objectives

The program objectives were based upon the needs assessment mentioned in the first step. The expected outcome of the implementation strategy was defined. What should be changed in the behaviour of the program participants (IPs) and what should be changed in their environment (the Institute)? Learning objectives for the IP were related to the following personal determinants: knowledge and skills, attitude, self-efficacy, and expectations. Change objectives were related to the following environmental determinants: availability, uniformity, and support. We connected the program objectives, the learning objectives and the change objectives. This approach enabled us



to define the concrete objectives, for which implementation could be developed. The main objective of the program was, to develop a strategy for the implementation of the guidelines for depression that suits the needs of the IPs.



**Figure 2.2:** Intervention Mapping Process, developed by Bartholomew et al [15]. Note: The term “insurance physician” was added to the original plan.

Semi-structured interviews held with 10 insurance physicians:

- 44 Questions about needs with regard to the medical content of the guidelines.
- 60 Questions about obstacles and facilitators when using the guidelines in practice.

Some examples of questions:

“Is the content of the guidelines for depression feasible in current practice?”

“How would you like to be trained in applying the guidelines?”

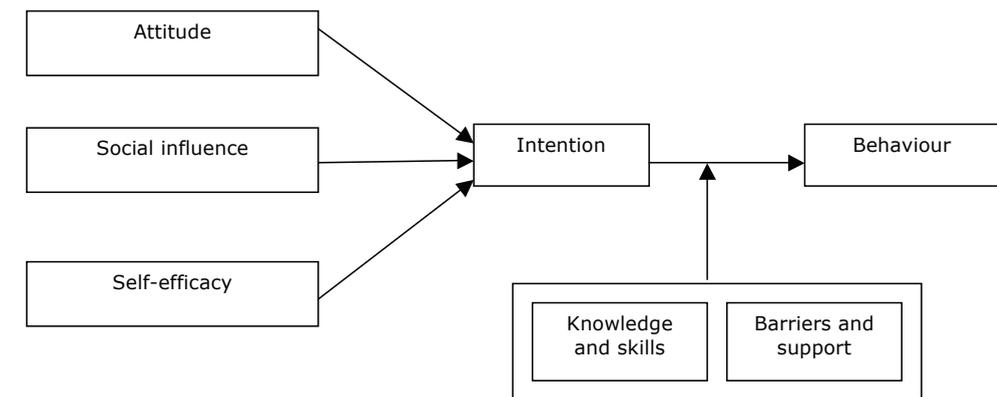
**Figure 2.3:** Needs assessment.

### Step 3: Selecting theory-based methods and practical strategies

In this step, suitable theory based methods and practical strategies were sought, in such a way that the chosen implementation reflects the scientific literature and evidence. These methods were subsequently translated into practical intervention strategies, the effectiveness of which already had been scientifically demonstrated. Learning objectives were defined for each of the personal determinants, and change objectives for each of the environmental determinants. The learning objectives, as laid down in the personal determinants of the IPs, can be achieved if the IPs are willing to change their behaviour. The barriers or the support in the process of guidelines implementation at the Institute can influence the change objectives for the environment. For this reason we looked for a theoretical model that describes behaviour, and how the environment influences behaviour. We used the Attitude, Social Influence and Self-Efficacy (ASE) model, derived from the Theory of Planned Behaviour (TPB) [22, 23]. The ASE model describes how a person’s attitude, social influence and self-efficacy (i.e. personal effectiveness) influence behaviour, as is shown in Figure 2.4 [24].

As used in this research setting, the ASE model may be explained as follows: the *behaviour* required from the IP consists of correct application of the guidelines for depression when assessing clients with depressive symptoms.

The *intention* to behave as described is determined by: 1) the IP’s *attitude* to the use of guidelines in general, and the guidelines for depression in particular; 2) the *social influence* exerted by the physician’s colleagues, and by the staff and managers who influence use of the guidelines, and who may influence availability and uniformity in using the guidelines; 3) the *self-efficacy* of the IP, or his/her confidence in his/her own ability to successfully apply the guidelines in practice [24]. The intention to use the guidelines does not necessarily result in their use in practice, i.e. the behaviour that is sought. The translation of intention into action is influenced by *barriers and support*, and by the existence of *knowledge and skills* (which can be increased by training) within the process and the organization. In the following steps we proceeded from theory to practical strategies; and from practical strategies to the intervention.



**Figure 2.4:** The ASE model, as defined by De Vries [24].

### Step 4: Program plan

The program plan for the implementation strategy was developed on the basis of the preceding steps: needs assessment, the matrix from step 2, and the theoretical model from step 3. Input for the program development process was obtained from semi-structured interviews and consultation rounds with 40 experts. These experts were mainly regular IPs, IPs responsible for appeal cases, (regional) staff physicians, and a few psychiatrists, training experts, and members of the management of the Institute. Preparing for these interviews and consultation rounds, we studied the disability assessment reports made by IPs to find out whether or not the main elements of the guidelines for depression could be found in the reports. Firstly, we had to look for the main elements of the guidelines for depression in the IP reports. Secondly, if we succeeded in finding them, these main elements could generate input for the training design. Finally, the main elements formed the basic assumption for the development of performance indicators (PI). In this investigation we screened IP reports on indicators of application of the guidelines for depression according to the saturation procedure. We screened reports until we reached the point, at which no new indicators for application of the guidelines could be found. Knowing that we could use the IP reports in our study, we designed a program plan, which incorporated the opinions of experts that we consulted. This program plan covered several aspects, such as PIs, instruments, training, and knowledge dissemination. In this stage we tried to match the implementation strategy with the needs and performance objectives of the IPs.

### Step 5: Planning the program implementation

After designing the program for the implementation strategy, we scanned the previous steps with a focus on objectives, methods and strategies, to ensure adoption by the IPs. A study of literature on the effects of implementation strategies was used to develop a suitable implementation strategy, consisting of instruments, training, testing and feedback

[25]. Given the context of the Institute, will the implementation strategy receive broad support, or could there be any obstacles in the implementation process? We consulted both users and stakeholders at the Institute regarding the content of the program and the implementation strategy. The users we consulted were the same 10 IPs who participated in the needs assessment. We then consulted six stakeholders at the Institute, i.e. the medical adviser, two regional staff physicians and three regional managers.

**Table 2.1:** *The needs of the insurance physicians with regard to guidelines for depression.*

Diagnosics	A list of the DSM IV criteria for depression and the DSM IV criteria for the most relevant differential diagnostic psychiatric disorders
Psychiatric examination	A list of psychiatric examination items on a desk mat
Seriousness depression	A method with which to determine the seriousness of depression in a uniform way. The Hamilton Rating Scale of Depression? (HRSD)
Seriousness and disability	Expert opinion to clarify the relationship between the seriousness of the disorder and the assessed disability
Prognosis	Need for evidence-based information about periods of recovery from depression in relation to treatment and co-morbidity
Guidelines for depression and other standards	Expert opinion on the relationship between the guidelines for depression and the standards: "Full disability entitlement on medical grounds" and "reduction in working hours" for partly disabled claimants
Coping styles	Information about personal characteristics and coping styles and how to distinguish between disease and behaviour

### Step 6: Evaluation

The intervention map can be used as an evaluation model for the development of the process, and for the effect of the corresponding intervention. In a future study we will evaluate the efficacy of a specific training in the implementation of the guidelines for depression in a two-armed RCT. The primary outcomes of the RCT will be the quality of the IP reports of the assessment of a claimant with depression, and the adherence of the IPs to the guidelines for depression. The outcomes of this RCT will be measured with performance indicators (PI) and questionnaires.



**Table 2.2:** *Implementation strategy: insurance physicians' wishes regarding educational training and support in the use of the guidelines.*

Training module	Form, implementation	Method
Introduction to the guidelines for depression	Experts from the curative sector and insurance physicians with knowledge of depression	Presentation of problems from curative and insurance-medical viewpoints; mutual questioning regarding experience and vision
Materials, tools	Summary card listing all diagnostic criteria	Practise in the use of the materials, and case histories
Case-histories	Group discussion and practise in applying the guidelines	What constitutes a good assessment? What is unclear? Why?
Work ability assessment	Insurance physician and psychiatrist/psychologist-researcher	Scientific insights, experiences, focus on problems (LFA)
Information on treatment possibilities	Experts from the curative sector	Current thinking on appropriate treatment. What questions can the insurance physician put to the curative physician?
Carrying out and interpreting psychiatric tests	Psychologist, psychiatrist	Different presentation in ethnic minorities (a high proportion of the claimants)
Detailed explanation and interpretation of the HRSD questionnaire	Psychologist, psychiatrist	Practise in the use of the questionnaire
Feedback	Insurance physician and guidelines author/researcher	Feedback from the profession; opportunity to ask questions

LFA=List of Functional Abilities; HRSD=Hamilton Rating Scale of Depression.

## Results

### Step 1: Needs assessment

Semi-structured interviews were held with 10 IPs. Almost all of these 10 IPs considered the guidelines to be useful as a reference, but indicated that they lacked information that is needed for direct use in practice. The specific items, representing the most important needs mentioned by the IPs with regard to the guidelines, are summarised in Table 2.1.

The IPs' wishes regarding implementation of the guidelines for depression were also established. They needed expert education. This should preferably be interactive training provided by experts, paying attention to practical relevance. The IPs wanted instructions on how to use the instruments, such as a desk mat listing all diagnostic



criteria, and psychiatric questionnaires based on case histories. In conclusion, the IPs wanted to be trained in applying the guidelines in practice with the help of experts and practical instruments. The IPs' wishes regarding the training module to support the guidelines are summarized in Table 2.2.

### Step 2: Program objectives

In this step we defined the behavioural and environmental determinants of the program, and translated them into performance objectives for the IPs and change objectives for the Institute. The IPs should learn how to use the guidelines for depression, and they should consider themselves capable of applying the guidelines in practice. By using and applying the guideline the IPs should believe that they could improve their performance with regard to their work ability assessments of claimants with depression. The Institute should increase the availability of the guidelines for the IPs, and should support the implementation by putting more emphasis on quality instead of productivity. Staff physicians should generate a strong influence in the use of the guidelines, by monitoring the IPs' reports on guideline adherence. The expected behaviour of the IPs is, that they will learn to apply the guidelines for depression. All the determinants of this behaviour were presented in a matrix (Table 2.3), crossed with the program objectives, showing the specifications of the program objectives for the IPs.

**Table 2.3:** Program objectives, learning objectives and change objectives.

Program objectives for insurance physician	Learning objectives for insurance physician (IP)'s personal determinants			Expectations	Change objectives for environmental determinants	
	Knowledge and skills	Attitude	Self-efficacy		Availability and uniformity	Support
IP makes thorough investigation and records findings transparently in the report.	IP has sufficient knowledge and skills to understand the guidelines and to implement it in practice.	IP accepts the guideline as a practical resource and a useful source of information.	IP considers him/herself capable of applying the guidelines in practice.	IP believes that use of the guidelines can make his/her examinations more thorough and transparent.	IP is trained to use the guideline and has the opportunity to practise using it during the training and subsequently in practice.	IP's quality-oriented activities are supported by National Institute for Employee Benefits Schemes by putting the emphasis on quality instead of productivity. Access to evidence-based medical info via Internet and/or library.
To do so, IP uses the guidelines in order to ensure quality and uniformity of the assessment	IP has the skills to perform the examination in line with applicable requirements.	IP supports the profession's general objective of fair assessment based on thoroughness, quality and uniformity.	IP considers him/herself capable of investigating issues associated with the assessment and obtaining guidance from the guideline, literature or colleagues.	IP believes that the quality and uniformity of his/her work ability assessments will be enhanced by the information in the guidelines.	Case histories are discussed amongst colleagues by reference to the guideline, enabling IP's to ask questions and learn from one another.	Staff of the Institute support IP in use of the guidelines and related activities. Staff physician encourages use of the guidelines, by testing the IP's reports on guideline adherence.
IP uses evidence-based information to support work ability assessment.	IP has sufficient evidence-based knowledge to recognize and address any lack of skills.	IP sees the guideline as a means to realizing the objective.		IP believes that the information in the guidelines will help him/her make more evidence-based work ability assessments.	Staff physicians provide all IP's with performance feedback and work with IP's to define individual learning programmes so that all attain a similar level.	Netherlands Association of Insurance Medicine supports IP's quality-oriented activities and encourages use of the guidelines.

IP=Insurance Physician.

### Step 3: Theory-based models and practical strategies

Practical interventions were chosen to realize the learning and change objectives mentioned in the Table 2.3. Subsequently, by putting the personal and environmental determinants in another matrix (Table 2.4) with the learning objectives, theory-based methods, and practical strategies, the required conditions for the development of the intervention were obtained. These methods and strategies were incorporated in the development of an interactive training with feedback. Adequate feedback on the performances of the IPs in the training should confirm their expectations that i.e. using the guidelines will contribute to more evidence-based assessments. IPs first must be aware of the guidelines, then become familiar with the guidelines, and finally believe that they are capable of working with the guidelines. IPs should be facilitated and stimulated by their environment in applying the guidelines, offering them training that suits to their needs. The Institute and the Netherlands Association for Insurance Medicine should support and involve the IPs in the development and implementation of guidelines. Determinants of learning and change objectives, and the associated strategies matched with theory-based methods are presented in Table 2.4.

### Step 4: Program plan

Research on the reports made by IPs when they assessed a claimant with depression, showed that these reports, indeed, did include the main elements of the guidelines for

depression. Saturation was achieved after 30 reports. Even without training IPs in the use of the guidelines, elements of the guidelines appeared in the IPs' reports. That made it possible to develop PIs for testing the reports for elements of the guidelines in the baseline situation. After this saturation procedure, we knew that we could use the IPs' reports to evaluate their implementation of the guidelines for depression. In addition, having found the main elements of the guidelines in the IPs' reports, we could determine the starting point for the design of the training. The planning of the implementation strategy was prepared and involved the following steps (see a, b and c below).

### a) Development of prototype instruments

The results of the interviews with the IPs were used in the development of the prototype instruments. With a view to aligning the instruments with the objectives of the guidelines, we consulted the adviser and secretary of the Health Council's Subcommittee on Depression. To supplement the guidelines for depression, a study was made of the literature on co-morbidity, prognostic risk factors, and the work capacity of individuals with depression. The result was a toolbox: a collection of instruments intended to facilitate application of the guidelines (see Table 2.5).

The desk mat showed on the front summarised information on the most essential points of the guidelines for depression. The back of the desk mat showed the relationship between the various relevant risk factors in a diagram based on the International Classification of Functioning, Disability and Health model (ICF model) [26], which was also used in the development of the insurance medicine guidelines. The ICF model is the framework within which the insurance physician operates when assessing the work

**Table 2.4:** Determinants of learning and change objectives and the associated strategies.

Determinant	Learning objectives for the insurance physician	Theory-based method	Practical strategy
Knowledge	Familiarity with the content of the guideline	Dissemination of training material Active learning from experts	Making guideline available in combination with practical instruments
Skills	The ability to apply knowledge in practice	Interactive group training	Interactive training in use of the guidelines
Attitude	Willingness to accept the guidelines and use them to improve quality	Persuasion by opinion leaders	Benefits highlighted during training and by staff and the Netherlands Association for Insurance Medicine
Self-efficacy	Belief in ability to use the guidelines in practice and finding answers to questions	Performance-related feedback	Positive individualised feedback during training and subsequently in practice, assistance with questions
Expectations	Expectation that the guideline will contribute to more evidence-based assessments	Individualized feedback and group performance audit data	Training in use of the guidelines with exercise case- histories, feedback at group and individual level
<b>Change objectives for the environment</b>			
Availability	The ability to practise, ask questions and work on personal performance	Feedback, personal improvement, planning	Practice in training, feedback on performance, support with questions
Uniformity	All insurance physicians covered by similar requirements	Quality-monitoring and quality-management	Staff physician appraises all insurance physicians using the same indicators
Support	Support from colleagues, staff, management and professional association, facilitation and, where necessary, amendment of the work process	In-built process reminders, quality management, support from opinion leaders	Quality evaluation by management, staff quality-oriented direction, promotion by the Netherlands Association for Insurance Medicine



ability of a disabled employee. Furthermore, a checklist contained items referring to the main points of the guidelines, such as the DSM IV criteria, seriousness of the depression, co-morbidity and treatment. When assessing a claimant with depression, the IPs can check all the relevant items and to make sure that they have not forgotten anything. Finally, the Hamilton Rating Scale for Depression (HRSD) [27] was added to the toolbox to assist the IP in the assessment of the severity of the depression at the time of the examination. Use of the HRSD needs to be included in the training for IPs in connection with implementation of the guidelines.

**Table 2.5:** Content of the toolbox.

Desk mat	Diagnosis and differential diagnosis based on the DSM-IV Assessment of the severity of depression Psychiatric examination Psychiatric co-morbidity Somatic co-morbidity Effective treatment methods Risk factors in relation to the severity and duration of disabilities The International Classification of Functioning, Disability and Health model (ICF model) [26] Key findings of the literature study referred to above
Checklist	Items referring to the main points of the guidelines for depression
HRSD [27]	Assessing the severity of depression

HRSD=Hamilton Rating Scale of Depression.



### **b) Refinement of the prototype instruments with help of a group of IP users**

The toolbox was shown to a group of users, consisting of 10 IPs. A questionnaire was used to establish their opinions with regard to the practicability, quality, content validity and added value of the instruments, as well as how easy they were to understand and the extent to which they allowed room for professional assessment. On the basis of the feedback from the user group, we searched for additional literature and adapted the instruments where necessary. This resulted in an amended and compressed desk mat and a check list. Also added to the toolbox was a consensus-based *list of the main abilities* that were thought to be associated with the work ability of employees with a major depressive disorder, and that could also be associated with the items of the HRSD [28].

### **c) Development of the training**

A separate group of experts, which included psychiatrists and training experts, was set up for consultations regarding the design of the training. This round of consultations resulted in the final training design as follows.

The IP should be given practical instructions about application of the guidelines for depression. This should include instructions on how to arrive at an evidence-based assessment of a depressive claimant's functional abilities, based on the knowledge presented in the guidelines. The learning objectives of the training appeared to be that the participating IPs trained their skills in making a diagnosis of depression, how to assess the severity of the depression and the disabilities, and how to report on the relationship between these issues. Meanwhile, they should learn how to give their assessment reports a solid base. To this end, the IPs should be provided with the aforementioned instruments. The training should start with a knowledge test based on the guidelines for depression. A psychiatrist who is familiar with the insurance-medical assessment system should then explain a number of important aspects of the assessment of depression on the basis of an interesting and recent case concerning an immigrant employee with an atypical presentation of depressive symptoms. Focus points should include diagnostics, the distinction between behaviour and disease, symptomatology, the relationships between symptoms and disabilities, assessment of the severity of the depression (including use of the HRSD), treatment, progression of the condition, and co-morbidity.

#### Instruments

- Summary of guideline on plastic desk mat
- Checklist referring to the guidelines
- Hamilton Rating Scale for Depression (HRSD)
- List of working abilities associated with depression

#### Training

- Interactive, feedback, based on realistic cases, experts as trainers

**Figure 2.5:** Program plan.

Subsequently, with a video recording of case study, an IP trainer should describe the practical aspects of using the instruments. The group of participants in the training should then be divided into subgroups, each focusing on a different part of the guidelines, to make assessments of the presented case. The relationship between the existing medical standards, "full disability entitlement on medical grounds", "reduction in working hours", and the guidelines for depression should be explained by the trainer. Different coping styles and personal characteristics of claimants should be integrated in the realistic cases presented during the training. In the training, the IPs should learn how to differentiate between the various types of coping styles of claimants with depression. Interactivity between the sub-groups and self-activation should alternate frequently, while feedback should be given by the trainer in an attempt to achieve the learning objectives for all the participants. When writing down their findings and conclusions, the participants should be instructed to use the essential elements of reasoning. Finally, the training day should end with an evaluation. In this kind of training design, the number of participants for each group should be limited to 20, because it is characterized by intensive communication, with feedback and interactivity between the participants and the trainer. The program plan is summarized in Figure 2.5.

### **Step 5: Program implementation**

We were interested in the opinions of experts, groups of users, and management and staff about implementation of the guidelines at the Institute, so that we could build up a picture of the context within which the IP works. The management and staff stated that, by implementing guidelines, they meet the requirements of the Ministry of Social Affairs. Furthermore, by implementing guidelines, the Institute might obtain more public support, and might face fewer complaints and appeals from claimants. Nevertheless, implementing guidelines could induce a loss of production. The IPs were pleased with the fact that, by carrying out research on the implementation of insurance medicine guidelines, attention will be paid to the quality and the content of their work. On the other hand, they realized that adopting guidelines might be a complex process for them, because they had to integrate working with the guidelines in their daily routine. The IPs had no previous history of working with guidelines. The IPs were in particular asked, to identify obstacles to and support for the use of guidelines for depression, and how the obstacles might be removed. One commonly identified obstacle to the use of a guideline was the emphasis placed on the quantity of the number of disability assessments to be made by an IP, which was imposed by the Institute. It was suggested that the Institute could facilitate the use of guidelines by placing more emphasis on quality, rather than quantity. Applying the guidelines thoroughly takes time, and productivity requirements limit the time that is available. The Institute was regarded as a productivity-driven organization. It was stated that staff physicians could stimulate the IPs to use the guidelines by giving them clear instructions about how to use them.

From the literature [29] and from consultations of decision-makers and implementers, it was found that the PIs for the guidelines can support the staff physicians in checking the IPs' reports on their adherence to the guideline. That would be a strong facilitator for using the guidelines according to the interviewed physicians. Furthermore the PIs could be used for feedback after training the IPs in the use of the guidelines, which was one of the needs of the IPs. By implementing guidelines, the decision-makers meet the requirements of the organization and the Ministry, but they might be faced with a loss of IP productivity. The IPs put more emphasis on quality by the implementation of guidelines, but wondered if they were capable enough of using the guidelines. The positive and negative features of the program implementation, as identified by the decision-makers, implementers and IPs, are summarized in Table 2.6.

**Table 2.6:** Positive and negative features of the program implementation for various parties concerned.

Parties involved	Positive features of program implementation	Negative features
<b>Decision-makers</b> Management of Socio-Medical Department	More public support Meets ministry requirements Fewer appeals and complaints	(Initial) loss of production Research takes time
<b>Implementers</b> (Regional) managers (Regional) staff physicians	Increased quality Fewer complaints Better-quality assessments More transparent decisions Easier test procedure to check reports	Loss of production, possibly temporary Appeals are not reduced Guidelines must not be rigid Legal status of guidelines: implication for appeals?
<b>Users</b> Insurance physicians	Useful guidelines and EBM information Guidelines with instruments tailored to IPs in practice Focus on quality and content Scope for professional assessment maintained	Learning a new approach takes time; integration in personal routine is an effort Stricter requirements made regarding examination and reporting. Will the extra workload be appraised and supported by staff and management? Legal status of guidelines: implication for appeals?
<b>Concerned</b> Claimants	More thorough and uniform claim assessment	Longer, more structured consultations (not necessarily a drawback)
<b>Researchers</b> Experts	Influence on content	Time input



## Step 6: Evaluation plan

The efficacy of the strategy for implementation of the guidelines, described in this article, will be compared to traditional implementation in a two-armed RCT. One group of IPs will receive specific training in applying the guidelines for depression, while the other group will continue with the traditional implementation of the guidelines. Hence, the specific training for the IPs will be the intervention in this RCT. Outcomes will be measured by PIs and questionnaires. The PIs measure the primary outcome, i.e. the behaviour of the IPs with regard to the guidelines, defined as: the quality of the IPs' reports of the assessment of a claimant with depression. The questionnaires not only measure the IPs' adherence to the guidelines, but also their satisfaction with the guidelines, which is a secondary outcome of this study. The questionnaires were developed on the basis of the literature and the ASE model [30, 31]. In the RCT the PIs and the questionnaires will determine the performance objectives of the IPs before and after the intervention. The process of the RCT and the specific training of the IPs in applying the guidelines will be evaluated in a process evaluation. The data for the process evaluation will be collected by means of specifically developed evaluation questionnaires. In order to illustrate how the effects of the implementation will be measured, the research model is presented in Figure 2.6.

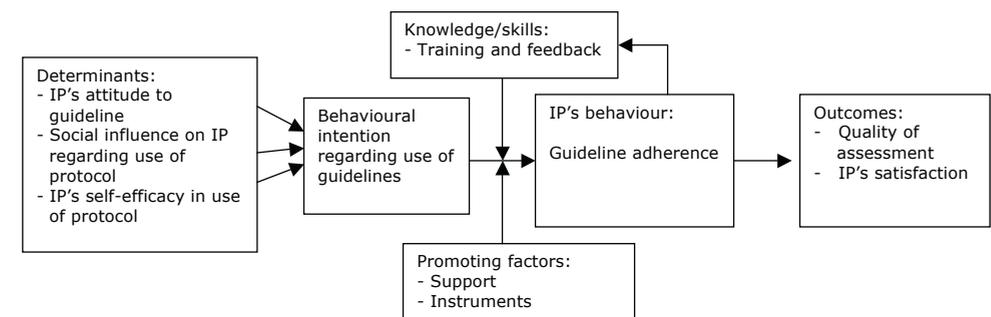


Figure 2.6: Research model, schematically represented on the basis of the ASE model. IP=Insurance Physician.

## Discussion

The aim of this study was to develop an implementation strategy to improve guideline adherence and the quality of the assessments made by IPs of the work ability of employees with depression. IM has its origin in public health, and in particular in health promotion programs. More recently IM has found its way into the field of occupational health medicine, where it has been used for Return to Work (RTW) interventions. The results of this study show that IM proved to be useful in the development of a strategy for the implementation of the insurance medicine guidelines for depression.



## Strengths and weaknesses

### Strengths

IM provides an implementation strategy framework in which a solid theoretical base and the participation of the IPs is integrated. The IPs will be more motivated to adopt the guidelines for depression if they are good compatible with daily practice and suit to their needs. By following all steps in the IM process, and with the help of 40 experts in the development of the instruments, performance indicators and training, we tried to achieve the practical feasibility of the guidelines for the IPs. We involved not only IPs and experts in the IM process, but also staff physicians, regional managers, the medical adviser and the top management of the Institute.

### Weaknesses

Generalization of the outcomes from IM studies might be difficult, because the IM process takes the local context into account. In our study, however the local context is set by the Institute: a national organization in which IPs assess the work abilities of claimants. Therefore, the outcomes of our study can only be generalized to other countries in which there is a central organization for employee benefits and IPs working with guidelines. Another weakness is that claimants with depression were not represented in this study. Nevertheless, we think, that a justifiable, careful and transparent assessment of work ability, in accordance with the guidelines, might be more acceptable for the claimant, than assessments without guidelines.

### Comparison with other studies

IM studies in insurance medicine are scarce, and only one has been published [20]. In that study, IM was used for the development of an RTW intervention program, whereas we used it for the development of a strategy for the implementation of the guidelines for depression by IPs. IM has been used as a systemic approach in designing a quality improvement intervention for general practitioners (GPs) [32]. In that study, using IM in the process of implementing guidelines for GPs, although time-consuming, appeared to be worthwhile. In Belgium, research has been carried out on the application of EBM and guidelines among IPs [33]. In that study, the IPs' knowledge about EBM and practical guidelines was found to be rather poor. Therefore, the authors recommended that high quality EBM and practical guidelines should be structured in such a way that they are useful for IPs. In our study we tried to achieve that aim with the added value of using IM. We tried to meet to the needs of the IPs, and we integrated EBM in the development of the instruments. This approach resulted in a tailor-made intervention: educational training for IPs in applying the guidelines for depression. However, with or without training, the application of guidelines by physicians remains a complex process, lacking in-depth knowledge about which factors are decisive in that process [13].

Integrated in the third step of the IM process, the ASE model, derived from the Theory of Planned Behaviour (TPB) [22, 24] appeared to be suitable to cover those factors. The adherence of physicians to the guidelines has been related to TPB in several studies [10, 34-36], and the overall conclusion was that health behaviour theory can be useful for improving adherence to clinical practice guidelines. Cabana [34] reviewed 76 studies on barriers to guideline adherence among physicians. From his review he compiled a list of barriers in physician adherence to guidelines. In our study we tried to overcome barriers in the adherence of physicians to guidelines by using IM for the development of our implementation strategy. The IPs will be made familiar with the guidelines for depression by a specific training. The guidelines were made more accessible for the IPs with the help of practical instruments. Bearing in mind the recommendations made by Grol in a review [37], we provided the IPs who participated in our study with a well-designed and well-prepared program for implementing the guidelines. In another review focussing on physicians' attitudes to guidelines [38], the authors stated that high satisfaction with guidelines does not necessarily results in practice changes. Individual physicians would not make significant changes without the necessary educational, organizational and structural changes in the health care system [38]. By using IM we tried to encourage the IPs to use the guidelines, taking into account all the aspects of behavioural change mentioned above.

### Practical relevance

The IM method cannot only be applied for implementing the guidelines for depression, but also for other insurance medicine guidelines at the Institute, and for guidelines in other disciplines outside the Institute. We expect that by using IM to develop a strategy for the implementation of insurance medicine guidelines, adherence of the IPs to the guidelines will improve. The educational training, as developed for the guidelines for depression, could be adjusted and prepared for the implementation of other insurance medicine guidelines. The implementation of the guidelines and the development of the PIs, has made quality testing possible. Auditing professional quality is a challenging issue and the social and professional need to measure quality has increased considerably in recent years. Occupational health processes have long been audited by means of indicators [10, 39, 40], and now indicators will be introduced into the field of insurance medicine to monitor IPs' assessments of the work ability of claimants. Transparency of professional decision-making can provide a basis for quality improvement and our study design is consistent with this trend of auditing quality improvement.

Further research is recommended to determine, whether an IM based strategy for the implementation of insurance medicine guidelines actually contributes to IP adherence to guidelines. We expect that the results of the RCT and the process evaluation will provide us with an answer to that question.



## Conclusions

This article describes the use of IM in the development of a strategy for the implementation of the insurance medicine guidelines for depression.

Although the implementation strategy we developed has yet to be evaluated, we may already conclude that the use of IM made it possible to develop guideline support instruments that are tailored to insurance medical practice. The instruments and PIs that were developed meet the needs of IPs, and take into account the context in which they will be used.

## Competing interests

The author(s) declare that they have no competing interests.

## Author's contributions

FZ wrote the manuscript. AJMS, KCR and JRA contributed to the manuscript. AJMS and JRA designed the study. AJvdB commented on the manuscript and will act as guarantor of this study. All authors have read and approved the final version of the manuscript.

## Acknowledgements

We would like to thank all the participating professionals, staff and decision-makers of the Institute, and the trainers and researchers for their contribution to the development of a strategy for the implementation of the insurance medicine guidelines for depression. The authors also thank K.M. van Beurden, MSc, for her practical assistance. This study was financially supported by the Dutch Institute for Employee Benefits Schemes.

## References

1. Brown P, Pirkis J: Mental health quality and outcome measurement and improvement in Australia. *Curr Opin Psychiatry* 2009, **22**:610-8.
2. Carek PJ, Dickerson LM, Boggan H, Diaz V: A limited effect on performance indicators from resident-initiated chart audits and clinical guideline education. *Fam Med* 2009, **41**:249-54.
3. Jones DW, Peterson ED, Bonow RO, Masoudi FA, Fonarow GC, Smith SCJ, et al.: Translating research into practice for healthcare providers: the American Heart Association's strategy for building healthier lives, free of cardiovascular diseases and stroke. *Circulation* 2008, **118**:687-96.
4. Crampton P, Perera R, Crengle S, Dowell A, Howden-Chapman P, Kearns R, et al.: What makes a good performance indicator? Devising primary care performance indicators for New Zealand. *N Z Med J* 2004, **117**:U820.

5. Martens JD, Van der Weijden T, Severens JL, De Clercq PA, De Bruijn DP, Kester ADM, et al.: The effect of computer reminders on GPs' prescribing behaviour: a cluster-randomised trial. *Int J Med Inform* 2007, **76**:S403-S416.
6. Nieuwenhuijsen K, Verbeek JHAM, De Boer AGEM, Blonk RWB, Van Dijk FJH: Validation of performance indicators for rehabilitation of workers with mental health problems. *Med Care* 2005, **43**:1034-42.
7. Smits PB, De Buissonje CD, Verbeek JH, Van Dijk FJ, Metz JC, Ten Cate OJ: Problem-based learning versus lecture-based learning in postgraduate medical education. *Scand J Work Environ Health* 2003, **29**:280-7.
8. Wobrock T, Weinmann S, Falkai P, Gaebel W: Quality assurance in psychiatry: quality indicators and guideline implementation. *Eur Arch Psychiatry Clin Neurosci* 2009, **259**:S219-S226.
9. Smolders M, Laurant M, Verhaak P, Prins M, Van Marwijk H, Penninx B, et al.: Which Physician and Practice Characteristics are Associated With Adherence to Evidence-Based Guidelines for Depressive and Anxiety Disorders? *Med Care* 2010, **48**:240-8.
10. Houghton G, Rouse A: Are NHS primary care performance indicator scores acceptable as markers of general practitioner quality? *Br J Gen Pract* 2004, **54**:341-4.
11. Van Amstel RJ, Anema JR, Jettinghoff K, Verbeek JH, Nauta AP, Van Putten DJ: [Limited change in the quality of the social medical guidance and in the satisfaction of sick-listed patients, after collaborative projects between general practitioners and occupational physicians]. *Ned Tijdschr Geneeskde* 2005, **149**:2407-12.
12. Van der Weide WE, Verbeek JH, Van Dijk FJ. Relation between indicators for quality of occupational rehabilitation of employees with low back pain. *Occup Environ Med* 1999, **56**:488-93.
13. Advani A, Shahar Y, Musen MA. Medical quality assessment by scoring adherence to guideline intentions. *Proc AMIA Symp* 2001:2-6.
14. Carek PJ, Dickerson LM, Boggan H, Diaz V: A limited effect on performance indicators from resident-initiated chart audits and clinical guideline education. *Fam Med* 2009, **41**:249-54.
15. O'Malley AS, Clancy C, Thompson J, Korabathina R, Meyer GS: Clinical practice guidelines and performance indicators as related--but often misunderstood--tools. *Jt Comm J Qual Saf* 2004, **30**:163-71.
16. Rebergen D, Hoenen J, Heinemans A, Bruinvels D, Bakker A, Van Mechelen W: Adherence to mental health guidelines by Dutch occupational physicians. *Occup Med* 2006, **56**:461-8.
17. Smolders M, Laurant M, Verhaak P, Prins M, Van Marwijk H, Penninx B, et al.: Adherence to evidence-based guidelines for depression and anxiety disorders is associated with recording of the diagnosis. *Gen Hosp Psychiatry* 2009, **31**:460-9.
18. Wilson APR, Gibbons C, Reeves BC, Hodgson B, Liu M, Plummer D, et al.: Surgical wound infection as a performance indicator: agreement of common definitions of wound infection in 4773 patients. *BMJ* 2004, **329**:720.
19. Zwerver F, Schellart AJM, Anema JR, Rammeloo K, Van der Beek AJ: Intervention mapping for the development of a strategy to implement the insurance medicine guidelines for depression. *BMC Public Health* 2011, **11**:9.
20. Bijl RV, Ravelli A, Van Zessen G: Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Soc Psychiatry Psychiatr Epidemiol* 1998, **33**:587-95.
21. Levav I, Rutz W: The WHO World Health Report 2001 new understanding--new hope. *Isr J Psychiatry Relat Sci* 2002, **39**:50-6.
22. Ormel J, Petukhova M, Chatterji S, Guilar-Gaxiola S, Alonso J, Angermeyer MC, et al.: Disability and treatment of specific mental and physical disorders across the world. *Br J Psychiatry* 2008, **192**:368-75.



23. Wang F, Lee HP, Lu C. Thermal-mechanical study of functionally graded dental implants with the finite element method. *J Biomed Mater Res A* 2007, **80**:146-58.
24. Health Council: *Insurance medicine guidelines for depression*. Den Haag: Gezondheidsraad, 2006.
25. Shoukri MM, Asyali MH, Donner A: Sample size requirements for the design of reliability study: review and new results. *Stat Methods Med Res* 2004, **13**:251-271.
26. SPSS Inc: *SPSS 15.0 Command Syntax Reference*, 2006.
27. Molenberghs G, Laenen A, Vangeneugden T: Estimating reliability and generalizability from hierarchical biomedical data. *J Biopharm Stat* 2007, **17**:595-627.
28. Shavelson RJ, Webb NM, Rowley GL: Generalizability theory. *American Psychologist* 1989, **44**:922-32.
29. Shavelson RJ, Webb NM: *Generalizability theory: a primer*. Newbury, Park, London, New Delhi: Sage Publications, 1991.
30. Euser AM, Le Cessie S, Finken MJ, Wit JM, Dekker FW: Reliability studies can be designed more efficiently by using variance components estimates from different sources. *J Clin Epidemiol* 2007, **60**:1010-1014.
31. Hamilton M: A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960, **23**:56-62.
32. De Boer WEL, Wind H, Van Dijk FJH, Willems HHB: Interviews for the assessment of long-term incapacity for work: a study on adherence to protocols and principles. *BMC Public Health* 2009, **9**:169.
33. Van Rijssen HJ, Schellart AJM, Anema JR, Van der Beek AJ. A theoretical framework to describe communication processes during medical disability assessment interviews. *BMC Public Health* 2009, **9**:375.
34. Hepner KA, Rowe M, Rost K, Hickey SC, Sherbourne CD, Ford DE, et al.: The effect of adherence to practice guidelines on depression outcomes. *Ann Intern Med* 2007, **147**:320-9.
35. Verbeek J, Spelten E, Kammeijer M, Sprangers M: Return to work of cancer survivors: a prospective cohort study into the quality of rehabilitation by occupational physicians. *Occup Environ Med* 2003, **60**:352-7.
36. Hermanides HS, Hulscher MEJL, Schouten JA, Prins JM, Geerlings SE: Development of quality indicators for the antibiotic treatment of complicated urinary tract infections: a first step to measure and improve care. *Clin Infect Dis* 2008, **46**:703-11.
37. Spreeuwiers D, De Boer AGE, Verbeek JHA, Van Beurden MM, Van Dijk FJH: Diagnosing and reporting of occupational diseases: a quality improvement study. *Occup Med* 2008, **58**:115-21.
38. Schellart AJM, Mulders HPG, Steenbeek R, Anema JR, Kroneman H, Besseling JMJ: Inter-doctor variations in the assessment of functional incapacities by insurance physicians. *BMC Public Health* 2011, **11**:864.
39. De Boer WEL, Besseling JMJ, Willems HHB: Organisation of disability evaluation in 15 countries. *Prat Organ Soins* 2007, **38**:205-17.

# Chapter

---

# 3

## **Development and reliability of performance indicators for measuring adherence to a guideline for depression by insurance physicians**

Disabil Rehabil 2011, 33:2535-43

**AJM Schellart  
F Zwerver  
DL Knol  
JR Anema  
AJ van der Beek**

## Abstract

**Introduction:** We wanted to measure adherence to the guideline for depression in disability assessments. The research questions we addressed were: How can we develop performance indicators (PIs) for adherence to the Dutch guideline for disability assessment of patients with depression and how can we measure the quality of the scores? What is the inter-rater reliability of these PIs? What is the quality of the PI scores?

**Methods:** PIs, developed by the researchers, were reviewed on various aspects, by a panel of seven experts in several consulting rounds. After adjustments, senior insurance physicians (IPs) attended two training sessions and scored the PIs on 10 different simulated case reports. Two researchers developed proxy 'gold standard' scores for these 10 case reports. To assess the inter-rater reliability and the quality of the scores, we calculated the intra-class correlations (ICC) and 95% confidence intervals (CI) of the PI scores and of the PI scores compared to the proxy 'gold standard', respectively.

**Results:** Six specific and relevant PIs resulted from the consultation of the panel of experts. The PI scores for the 10 case reports, rated by seven (of the eight) senior IPs who completed both training sessions, showed that the PIs were not reliable at individual level (ICC = 0.543; 95% CI 0.426-0.642). However, the ICC became more reliable as an average of two raters was calculated (ICC = 0.704). The ICC of the PI scores with the proxy 'gold standard' was 0.538 (95% CI 0.419-0.640), but the quality was higher when calculated as an average of two raters (ICC = 0.700).

**Conclusion:** The PIs for adherence to the guideline were sufficiently reliable, and the quality of their scores was adequate if at least two well-trained raters were involved. The senior IPs evaluated the feasibility of the PIs as good, with a prerequisite of sufficient training. This method may be interesting for measuring guideline adherence and quality of disability assessments in general.



## Introduction

### Background

Worldwide, evidence-based guidelines are being implemented in an attempt to improve health care [1-3]. Translating scientific evidence into practical guidelines to help health care professionals to improve the quality of care is an important challenge. Most of the research concerning guidelines has focused on the questions: to what extent are guidelines being applied in medical practice, and how is the application of guidelines related to patient outcomes, such as quality of care? Thus, the quality of the assessment of provided health care is a core problem in guideline-based care. Performance indicators (PIs) are frequently used to assess the quality of the care that is provided, given a certain outcome. PIs have been defined as measurable elements of practice performance, for which there is evidence or consensus that they can be used to assess quality and hence, also a change in the quality of care that is provided [4]. Assessment of the quality of provided health care, and development of the accompanying PIs, has attracted interest in different fields of health care: clinical, primary and occupational care [5-8].

Guidelines and PIs support the mission of the American Heart Association "to build healthier lives, free of cardiovascular diseases and stroke", by translating evidence into practice for health care professionals [3]. In a German study of guideline implementation in psychiatry [8], it was stated that the routine mental health care did not correspond to the standards of the medical profession itself. PIs and improved adherence to guidelines could improve the sub optimal outcomes, and narrow the gap between optimal and routine care in psychiatry. A primary care study showed that only a minority of the patients with a depressive or anxiety disorder was treated in accordance with the guidelines, and the reasons for this conclusion were attributed to physician characteristics, and not to practice characteristics [9]. In another primary care study, it appeared that general practitioners tend to find externally imposed measurements irrelevant [10]. These general practitioners probably experience such measurements as an infringement of their autonomy. Therefore, PIs should be relevant and acceptable for practicing physicians. In his article "What makes a good performance indicator", Crampton stated that PIs should reflect important aspects of health status, be attributable to health care, be linked to health outcomes, be sensitive to change, be based on reliable and valid information, be precisely defined, and be easily quantifiable [4].

In occupational health care, the quality of the care has been studied on the basis of several guidelines, such as the guideline for low back pain, and the guideline for mental illness [6, 11, 12]. The outcomes were the occupational physicians' adherence to the guidelines and return to work of the employee. In the literature, the most frequently mentioned conclusion about quality assessment and guidelines in any field of health care is: the performance of the professional leaves room for improvement. Improvement in the quality of care could be based on retrospectively comparing the professionals'



actions to the recommendations in the guidelines [13]. This indicates the significance of establishing and measuring the performance of health care professionals.

PIs can be developed in different forms or structures. Most PIs consist of a list of criteria derived from guidelines or other evidence-based processes [6, 12, 14-17]. However, different forms of PIs are possible, depending on different types of outcomes. PIs are used to measure the quality of care in different disciplines, in primary and secondary care, and in the public health setting in general [18].

Guidelines have recently been introduced in the field of insurance medicine in the Netherlands. Since 2006, 19 disease specific guidelines have been developed by the Dutch Health Council, to be used by insurance physicians (IPs), with the intention to promote evidence-based medicine in daily practice. By implementing guidelines, the quality and transparency of disability assessments made by IPs should increase [19]. PIs should account for the complexity of decisions made by the IPs, assessing the work disability of a claimant according to the guideline. We selected the guideline for depression from the 19 insurance medicine guidelines, because depression is an ever increasing contributor to work disability [20-23]. The guidelines for depression contain not only evidence-based information concerning epidemiology, diagnostic aspects, such as DSM IV classification and relevant co-morbidity, treatment and therapy of the disorder, but also provides guidelines for the IP, as to how to finally integrate these findings in the assessment of the limitations in working ability [24]. For example; an IP should describe reasons for stagnation in recovery of the disorder and return to work of a client by assessing the severity of the symptoms, the presence of co-morbidity, the risk factors that sustain the symptoms, and the clients' compliance with therapy, which all together define the course and prognosis of the disorder and eventually return to work. We divided the guidelines into 6 main performance goals, and developed PIs for each of these points. Therefore, the developed PIs had to cover these main points of the guidelines for depression, and should be valid, reliable and feasible. In insurance medicine, the performance of the IPs is reflected in their disability reports, and for reasons of applicability the PIs have to be appropriate for these reports.

### Research questions

The aim of this study was to develop PIs for scoring PI adherence to the guideline for depression by IPs and to measure the quality of the scores obtained from the PIs. Furthermore, we aimed to assess the inter-rater reliability of these indicators and the quality of the scores. The research questions we addressed were:

- a) How can we measure the adherence to the guideline for depression by means of a scoring method based on PIs and b) how can we measure the quality of the scores?
- b) If simulated case-reports are scored, a) what is the inter-rater reliability of these PIs at individual and group level, and b) what is the quality of the PI scores.

## Method

### Regarding the first research question

Three researchers (among whom two IPs) developed a first version of 10 PIs. These PIs were in the form of decision trees, based on the content of the guideline for depression. For these 10 PIs we consulted 7 expert IPs, who were familiar with the guidelines for depression and who had expertise in the field of insurance medicine practice. Among these 7 expert IPs were staff IPs, who are used to judge other IPs' case-reports, trainer IPs, as well as IPs who participated in the work group which developed the guidelines for depression. These 7 expert IPs discussed this first version with regard to content, specificity, applicability and relevance in practice. Based on the opinions of these experts, the 10 PIs were adjusted within the first two rounds. In a third and final round, they were checked for content validity with the guideline for depression (FZ and JRA) and for consistency (AJMS). This resulted in PIs, which focussed on the main aspects of the guidelines for depression.

For testing the PIs 10 case-reports, each three to four pages long, were written by one researcher (FZ). These case-reports were constructed on base of IPs' disability assessment reports concerning various clients with depression, which were assessed in practice. These assessment reports were made anonymously and edited by one of the researchers (FZ) for the purpose of testing the PIs. That resulted in 10 edited different depression case-reports, some of which completely met the guidelines for depression (implying that all PIs should end up into an Adequate score), while other case-reports had built in imperfections. For example, in a complete case-report only the remarks about the severity of the disorder were deliberately left out. The PIs applied to this case-report, should detect this degree of imperfection in that case-report, by resulting in all Adequate scores, apart from the score for PI-2 that measures the severity of the disorder. These case-reports had an identical format, i.e. 9 sub-divisions: background, questions and purpose, examinations (from medical file and medical examination), diagnosis, summary and evaluation of findings, prognosis, patient's reaction, conclusion, and planning. The edited case-reports were aimed to have the same degree of difficulty.

Eight Test IPs, who were different from the expert IPs, were invited for a one-day training, in which they also tested the PIs on the 10 simulated case-reports. These 8 Test IPs were recruited from the researchers' network on base of their affinity to guidelines in general and their thorough knowledge of the guidelines for depression and their conscientious way of working in particular. The Test IPs were not involved in the development of the PIs in any way. The training took place in two groups of four participants to increase the possibility of interaction during the training. The same two trainers (FZ and JRA) trained both groups. Before the training, the 8 Test IPs received the guidelines for depression as a booklet and a manual on to how to apply the PIs to the case-reports of claimants with a depression in general. An eleventh case-report was

written before and, which was used as an example for scoring the PIs, in an interactive way, during the training. The first part of the training consisted of an explanation of the PIs, and interactive scoring of the example case-report. In the second part of the training, the 8 Test IPs scored the PIs for the 10 written case-reports individually. We instructed the Test IPs to score conservatively. The sequence of the 10 case-reports was at random. After the last case-reports had been scored there was a group discussion on the acceptability and feasibility of using of the PIs on the 10 case-reports. We also asked the Test IPs about their opinion of the training and the feasibility of the PIs in practice. They then handed over the written case-reports and the forms with the PI scores to the researchers. About 16 months after this first training, the same 8 Test IPs were invited to return again for a one-day refresher training session, with the same example case-report and manual, to score the same 10 case-reports in a random sequence. The reason for this second training session was that the start of the trial, for which the PIs together with the Test IPs formed the measurement instrument, had been delayed for that period. To keep the Test IPs' knowledge of the PIs and the guidelines at the required level for the measurements, we had to plan this refresher training. The output of both training sessions, i.e. the PI scores on the 10 case-reports, were used for the calculations of the inter-rater reliability between the Test IPs.

To calculate a proxy score for a "gold standard", two researchers with experience in the field of depression (FZ and JRA) independently, and blinded for the results of the Test IPs, scored the PIs for each of the 10 case-reports. In a consensus meeting the differences in scores were discussed. Most of the differences appeared to be due to differences in the scoring of multi-interpretable argumentations in the case-reports. During the consensus meeting it was agreed to score "conservatively", i.e. to give a score "not adequate" in case of doubt.

### Regarding the second research question

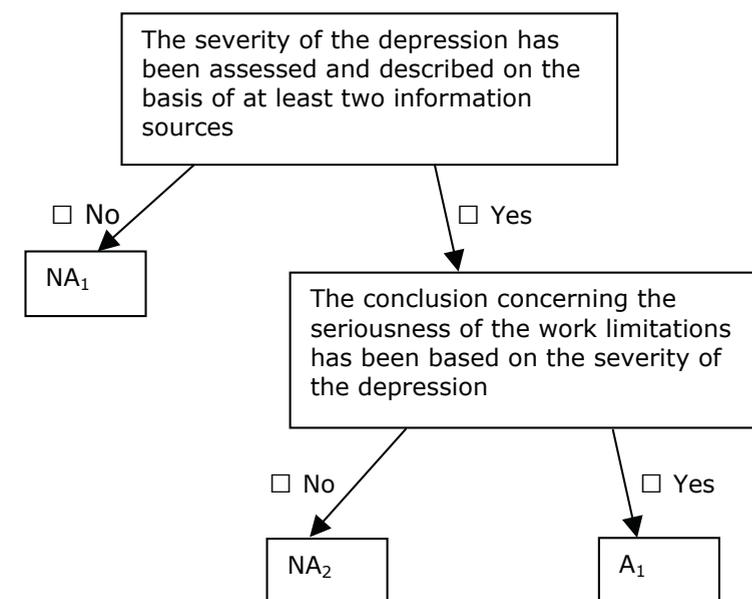
With 8 Test IPs and 10 case-reports (with 6 measurements each), the criteria were met to detect even a low inter-rater reliability and variance (estimated  $\rho \approx 0.12$ ; variance estimated  $\rho \approx 0.008$ ) [25].

We performed our analyses with linear mixed model, which provides the possibility of modelling variances (and covariances), as well as the possibility to account for hierarchical data [26]. Deriving the intraclass correlation (ICC) from these variances is straightforward [27].

We calculated the intra-class correlations (ICCs) for the inter-rater reliability of the scores of the Test IPs in the two training sessions separately. For this calculation we used a formula derived from the generalizability theory [28, 29], with the variance components of a linear mixed model [26, 27]: the case-reports, the PIs nested within the case-reports, the raters, the interaction between case-reports and raters, and the residual variance. The sum of the variance components for the case-reports and PIs



nested within case-reports formed the universe variance and the sum of the other variance components formed the absolute error variance in the ICC calculation. The ICC is defined by the ratio of universe variance and the sum of universe and absolute error variance. We considered an ICC of 0.70 as acceptable. The 95% confidence intervals (CIs) of the ICCs were calculated from the variance components, using the Fisher's Z transformation and the delta method described by Euser et al. [30].



**Figure 3.1:** Performance Indicator-2. A=adequate; NA=not adequate.

To measure the quality of the scores, we calculated for the two training sessions separately the ICC between the individual scores of the Test IPs and the proxy score for the "gold standard". Again, we used a linear mixed model, in which the IPs scores for the PIs was extended with the 60 proxy scores for the "gold standard", and an extra factor "type" was added to indicate whether we were dealing with the IPs scores or with proxy scores for the "gold standard". We estimated the variance components of this linear mixed model: the case-reports, the raters nested within the type and the PIs nested within case-reports, and the residual variance. Again, the sum of the variance components for the case-reports and the PIs nested within the case-reports formed the universe variance, and the sum of the other variance components formed the absolute error variance in the ICC. The 95% CIs were calculated in the same way as described before. All calculations were performed with SPSS 15.02.

## Results

The primary versions of the PIs were adjusted after rounds of consultations resulting in a decrease of the number of PIs, from 10 to 6 final PIs in the form of a decision tree. These 6 PIs also were made more manageable for the Test IPs to be used in practice, without losing their force in measuring the guidelines for depression. These PIs were focussed on the content of the disability assessment of a client with depression. Within the decision tree of a PI, one may score whether or not a certain step in the decision tree is made according to the guidelines in a case-report, ending up into an Adequate or Not adequate score. Performance Indicator-2, for example (see Figure 3.1), had two “not adequate” exits (NA1 and NA2) and one “adequate” exit (A1). This PI registered the information sources on the basis of which the severity of the depression was described. The IP may obtain the information from two sources: the medical file or the history-taking, and the medical examination, e.g. using the Hamilton Rating Scale for Depression [31]. Information from one of the two information sources was considered as adequate (“Yes”), i.e. according to the guideline. The next step was adequate (“Yes”), if the IP considered the severity of the depression as an argument for the seriousness of the limitations mentioned in the case-report. There is a relationship between the two. For example, in the case of a major depression, a reduction in working hours may be suggested because of a severely disturbed sleeping pattern and lack of energy.

**Table 3.1:** Number of exits per performance indicator.

Performance indicator (PI)	Number of exits “Not Adequate” (NA)	Number of exits “Adequate” (A)
PI-1	NA1 ... NA7	A1 ... A4
PI-2	NA1 ... NA2	A1
PI-3	NA1	A1
PI-4	NA1 ... NA4	A1 ... A2
PI-5	NA1 ... NA4	A1
PI-6	NA1 ... NA4	A1 ... A2

Table 3.1 shows the number of “non-adequate” (NA) and “adequate” (A) exits of the 6 PIs.



**Table 3.2:** Topics of performance indicators for the guideline for depression.

PI 1	Correct diagnosis <ul style="list-style-type: none"> <li>• DSM-IV criteria for depressive disorder</li> </ul>
PI 2	Determination of severity of the disorder <ul style="list-style-type: none"> <li>• Source: medical examination or e.g., information of curative physician, HRSD</li> <li>• Relation between severity of the disorder and the limitations</li> </ul>
PI 3	Origin, course and prognosis of the disorder <ul style="list-style-type: none"> <li>• Risk factors for depressive disorder</li> <li>• Course of depressive disorder</li> <li>• Substantiated prognosis of depressive disorder</li> </ul>
PI 4	Co-morbidity <ul style="list-style-type: none"> <li>• Presence or absence of co-morbidity</li> <li>• Influence of co-morbidity on prognosis and limitations</li> </ul>
PI 5	Evaluation of care and cure <ul style="list-style-type: none"> <li>• Level of information about claimant and medical treatment</li> <li>• Action for required information if necessary</li> <li>• Reasons for stagnation in recovery of functioning</li> <li>• Medical treatment related to rehabilitation</li> </ul>
PI 6	Assessment of work limitations <ul style="list-style-type: none"> <li>• Work limitations related to the severity of depressive disorder</li> <li>• Work limitations substantiated to insurance medicine standards</li> </ul>

The topics of the 6 PIs (see Table 3.2) were logically connected. For example, on the basis of the diagnosis of depression (PI-1), the severity of this depression as argument for the related work limitations (PI-2), together with a bad prognosis (PI-3), despite the absence of co-morbidity (PI-4) and an adequate care and cure provided by the curative physicians (PI-5), the IP may decide from the case-report that the patient has certain limitations in working ability and restrictions in the number of working hours (PI-6).

The PIs are ending up into different possible scores at all branches of all 6 decision trees varying from A1 to A4 and from NA1 to NA7. For example, if in a case-report written by an IP the score on PI-1 is NA4, then this PI score implicates that the IP who wrote this case-report had reported some things right, i.e. the first branches, but finally failed in the fourth branch to reach an Adequate score. Consequently, the NA4 score is closer to the guidelines than the NA1 score. If a case-report has an NA1 score on PI-1, then the IP who wrote this case-report went wrong immediately at the first branch. Hence, there exists a certain order in the range of PI scores, implying variation in distance to the guidelines for depression. In this example, the first IP has a better performance according to the guidelines than the second IP on PI-1.

**Table 3.3:** Proxy scores for the “gold standard”.

Case-report number*	Adequate	Not Adequate
Case 1	5	1
Case 2	4	2
Case 3	5	1
Case 4	5	1
Case 5	4	2
Case 6	4	2
Case 7	6	0
Case 8	4	2
Case 9	1	5
Case 10	6	0
Total	44	16

\*Sequence conform the first training.

Taking into account the distance of the NAs and As within the PIs to the guidelines for depression, and the differences across the PIs, the possible PI scores were recoded in the following way to form a scale: (NA1=1), (NA2=2), (NA3=3), (NA4..NA7=4), (A1=5), (A2=6), (A3..A4=7). This scale was formed on base of all possible PI scores in order from highest to lowest match with the guidelines for depression.

**Table 3.4:** Variance components\* of scores for 6 performance indicators for 10 case-reports scored by senior insurance physicians (IPs).

Effect	First training, 8 IPs, incl rater 4, (n=480)		Second training, 7 IPs, excl rater 4, (n=420)	
	Variance components	Standard error	Variance components	Standard error
Case-reports	0.000	0.000	0.000	0.000
PI(case-report)	1.288	0.273	1.763	0.363
Raters	0.051	0.046	0.042	0.039
Case-reports*raters	0.105	0.065	0.007	0.051
Residual	1.478	0.112	1.436	0.117

\*Estimated with linear mixed model.

The proxy score for the “gold standard” consisted of 60 PI scores. 44 PI scores for the 10 case-reports were considered as “adequate”, 8 case-reports were given 16 “not



adequate” scores for various PIs, and 2 case-reports were scored as perfect adherence to the guideline for depression (see Table 3.3).

Of the eight selected Test IPs, five were men and three were women, their ages ranged between 40 and 63 years, and they all had more than 10 years of experience as an IP. They all completed the first training, and except for one senior IP who retired in the meantime (rater 4), they all completed the second training. Table 3.4 shows the variance components, as estimated with the linear mixed model for the first training with 8 IPs (including rater 4) and the second training with 7 IPs (excluding rater 4).

Table 3.5 presents the ICCs for the first and the second training, using a different number of raters. The ICCs for one rater were calculated from the variance components given in Table 3.5. The ICCs for two or more raters were derived from the single ICCs, using Shavelson’s formula [28, 29]. The ICC in the second training (excluding rater 4) was only a slightly higher than the ICC in the first training, but both were a lot lower than 0.70. The ICCs as an average of more raters were equal or higher than 0.70 for three or more raters in the first training and for two or more raters in the second training.

**Table 3.5:** Intra-Class Correlations (95% CI)\* of scores of 6 performance indicators for 10 case-reports scored by senior insurance physicians (IPs).

Number of raters k	First training, 8 IPs, incl rater 4, (n=480)	Second training, 7 IPs, excl rater 4, (n=420)
1 rater	0.44 (0.33-0.54)	0.54 (0.43-0.64)
2 raters	0.61 (0.50-0.71)	0.70 (0.60-0.78)
3 raters	0.70 (0.60-0.78)	0.78 (0.69-0.85)
4 raters	0.76 (0.67-0.83)	0.83 (0.75-0.88)
5 raters	0.80 (0.71-0.86)	0.86 (0.79-0.90)
6 raters	0.83 (0.75-0.88)	0.87 (0.82-0.92)
7 raters	0.85 (0.78-0.90)	0.89 (0.84-0.93)
8 raters	0.86 (0.80-0.91)	0.91 (0.86-0.94)

\*Intra Class Correlations (ICCs) calculated on the basis of the single ICCs (i.e. one rater) with the formula:  $ICC / [(ICC + (1 - ICC) / k)]$ , where k is the number of raters. Confidence intervals (95% CI) calculated from variance components using Fisher’s Z transformation and the delta method.

The ICCs (95% CI) between the IP scores (1 ... 7) and the proxy score for the “gold standard” were calculated from the variance components of the linear mixed model (not shown here). For the first training (including rater 4) the ICC was 0.456 (95% CI 0.342-0.556), and for the second training (excluding rater 4) it was 0.538 (95% CI 0.419-0.640). The ICC of the second training was only a slightly higher than the ICC of the first training. Again, the ICCs were equal or higher than 0.700 for two or more raters in the second training.

## Discussion

### Main findings

One important result is that it is possible to develop PIs that represent the key aspects of the guideline for depression. The PIs were developed on base of these guidelines and were adjusted by both experts and researchers, keeping the guidelines in mind. Specially trained Test IPs judged the PIs as feasible in practice.

However, scoring the PIs for the guideline for depression can not be done reliably by one rater, because the resulting ICC would be too low ( $<0.7$ ). However, groups of at least two well trained raters (Test IPs) obtained a more reliable score, with an ICC of 0.700 or higher, when measuring the adherence of a group of IPs to the guideline for depression. Given our proxy scores for a “gold standard” of the 10 case-reports, the ICC of the scores as an average of two raters indicated that the scores of the Test IPs had adequate content validity. The participating senior IPs considered the PIs to be valuable and feasible in practice.

### Strengths and weaknesses

A strength of this study is that we developed PIs that have been directly derived from the guideline. These PIs were assessed by experts, and adapted according to their advice. This method of development ensured good content validity. The PIs assessed the case-reports on all aspects of the guideline and on the basis of the arguments used in the case-reports. Another strength is that we developed proxy scores for a “gold standard” for 10 simulated case-reports, which provided us with an indication of the content validity of the PI scores. The senior IPs we selected were positive about the training, and found the 10 simulated case-reports realistic.

A weakness is the relative complexity of the PIs, and the fact that they are guideline specific. PIs must be developed for individual guidelines, and this requires an intensive training. Another weakness is that, for practical reasons we were not able to develop a real “gold standard” for the scores of the 10 case-reports used in the training: inevitably there is a certain subjective element in scoring the PIs, but this subjective element is made inter-subjective by the scores.

Another point should also be discussed as well. A complete disability assessment comprises more than can be written down in a report. For instance, the assessment interview itself [32], i.e. the communication between the IP and the claimant [33], cannot be covered by our PIs. It may be argued that it is unsatisfactory to assess the quality of all the IP’s work by measuring only the quality of the IP’s reports. However, an assessment report with close adherence to the guideline according to the PIs, represents a disability assessment that meets the requirements of the guideline.



### Comparison with other studies

Quality assessment of the health care that is provided, by means of PIs is common in all fields of care, except for insurance medicine. We found no PIs at all for insurance medicine. Quality assessment with PIs is routine practise in clinical care, primary care and in occupational health care. The method of scoring of our PIs proved to be sufficiently reliable, valid, and well feasible. Furthermore, our PIs were constructed on the basis of branching logic, in the form of decision trees, providing detailed information with outcomes in a whole range of appropriate or inappropriate scores for the guideline. In other studies the PI scores were often dichotomised in appropriate or below standard relative to the guideline, leading to less precise adherence scores [6, 12, 14, 17, 34, 35]. The validity of the PIs is generally defined by means of the Delphi method or other forms of expert panels, as in our study [6, 17, 19, 34]. However, the reliability and feasibility of the PIs has seldom been tested before application in practice [36].

Once the PIs were developed and subsequently applied, there have sometimes been problems in interpreting PI scores, for different reasons: limitation in the interpretation of the quality scores because the quality of the score could depend on the content of the guideline [37]; relationship between PIs and outcomes remains unclear: return to work as outcome variable for occupational rehabilitation is influenced by too many variables [12]; registered adherence to guidelines was contrary to self-reported adherence to guidelines [16]; PIs were applied to cases that were too heterogeneous [35, 36]. We believe that in our study we can deal with the aforementioned problems for various reasons. Firstly, we investigated the match between the IPs’ reports and the guideline for depression in a saturation procedure, based on the grounded theory [19]. Only after finishing this procedure, did we start to develop the draft PIs based on the guideline and the structure of the IPs’ reports. Secondly, the performance of the IP, when assessing a claimant with a depressive disorder, is only influenced by the IP him/herself, and not by any other variables, such as the role of the employer in the outcome of return to work. Thirdly, our outcomes on adherence to the guideline and the quality of the IP’s reports are well defined in the guideline and in the set of PIs. Finally, we tested our PIs on 10 different simulated case-reports, which were based on real-life cases of claimants with a depressive disorder. The PIs were able to cover this case-mix.

### Practical relevance

Compared to usual disability assessment testing in the Netherlands, which is mainly based on procedural aspects, the PIs we developed mainly focussed on the content of the disability assessment, as reflected in the case-report. We think this is an important step in the stimulation of quality improvement, because the PI scores now provide feedback on the content of the assessment. PIs can detect weak points in the routine work of IPs, providing an opportunity for improvement. Moreover, to improve the quality of the assessments reports, feedback, whether at group or individual level, may be intended to

reduce systematic inter-doctor variations in work disability outcomes in practice [38], for which the IPs can use the PIs as a “script” when writing their case-reports.

The PIs were developed for the guideline for depression in the field of insurance medicine. However, the main framework of these PIs can also be adapted for other guidelines. We estimated that PIs can probably be developed for 12 of the 19 existing guidelines. It would be even if PIs were developed together with the guidelines, tested in practice and after adjustment, if necessary, implemented at the same time. Furthermore, the PIs can be used as an instrument to IP adherence to guidelines in general. The resulting feedback, when the IPs are confronted with the PI scores, may also support the maintenance of the guidelines.

We think that this method could be interesting for organisations and physicians in other countries, where physicians have to apply guidelines with regard to work disability assessments [39].

## Conclusions

PIs, in the form of decision trees, were found to be a reliable instrument if at least two trained raters are involved, to assess adherence to guidelines in the case-reports written by IPs for the work disability assessment of claimants with a depressive disorder. In this way, the quality of grouped disability assessments can be measured indirectly. These PI scores at group level can be used as starting point for individual feedback and discussion.

The PIs might be suitable for other insurance medicine guidelines, and could also be interesting for the measurement of guideline adherence and the quality of disability assessments in other countries.

## Contributors and conflict of interest

The authors declare that they participated in the study and made the following contributions to the study, and that they have seen and approved the final version. AJMS, FZ, JRA, DLK and AJvB contributed to the conception and design of this study. AJMS en DLK contributed to the analysis. AJMS and FZ wrote the manuscript. JRA, DLK and AJvB revised and commented on the manuscript. AJMS, JRA and AJvB will act as guarantors of this study. The corresponding author (AJvB) had full access to all the data in the study and had final responsibility for the decision to submit for publication. They declare that they have no competing interests. Funding sources: FZ, JRA and AJMS are (partially) funded by UWV. The study sponsor had no (decisive) role in the study design, in the collection, analysis, or interpretation of the data, in the writing of the case-reports, or in the decision to submit the paper for publication. The design of this study was laboratorial and for data collection fictitious but realistic case-reports were used. Consequently, the Medical Ethics Committee was not involved.



## Acknowledgements

The authors thank the senior insurance physicians who participated in this research. The Research Center for Insurance Medicine AMC-UMCG-UWV-VUmc, in Amsterdam, is a joint initiative of the Academic Medical Center (AMC), University Medical Center Groningen (UMCG), National Institute for Employment Benefits Schemes (UWV), and the VU University Medical Centre (VUmc).

## References

1. Brown P, Pirkis J: Mental health quality and outcome measurement and improvement in Australia. *Curr Opin Psychiatry* 2009, **22**:610-8.
2. Carek PJ, Dickerson LM, Boggan H, Diaz V: A limited effect on performance indicators from resident-initiated chart audits and clinical guideline education. *Fam Med* 2009, **41**:249-54.
3. Jones DW, Peterson ED, Bonow RO, Masoudi FA, Fonarow GC, Smith SCJ, et al.: Translating research into practice for healthcare providers: the American Heart Association's strategy for building healthier lives, free of cardiovascular diseases and stroke. *Circulation* 2008, **118**:687-96.
4. Crampton P, Perera R, Crengle S, Dowell A, Howden-Chapman P, Kearns R, et al.: What makes a good performance indicator? Devising primary care performance indicators for New Zealand. *N Z Med J* 2004, **117**:U820.
5. Martens JD, Van der Weijden T, Severens JL, De Clercq PA, De Bruijn DP, Kester ADM, et al.: The effect of computer reminders on GPs' prescribing behaviour: a cluster-randomised trial. *Int J Med Inform* 2007, **76**:S403-S416.
6. Nieuwenhuijsen K, Verbeek JHAM, De Boer AGEM, Blonk RWB, Van Dijk FJH: Validation of performance indicators for rehabilitation of workers with mental health problems. *Med Care* 2005, **43**:1034-42.
7. Smits PB, De Buissonje CD, Verbeek JH, Van Dijk FJ, Metz JC, Ten Cate OJ: Problem-based learning versus lecture-based learning in postgraduate medical education. *Scand J Work Environ Health* 2003, **29**:280-7.
8. Wobrock T, Weinmann S, Falkai P, Gaebel W: Quality assurance in psychiatry: quality indicators and guideline implementation. *Eur Arch Psychiatry Clin Neurosci* 2009, **259**:S219-S226.
9. Smolders M, Laurant M, Verhaak P, Prins M, Van Marwijk H, Penninx B, et al.: Which Physician and Practice Characteristics are Associated With Adherence to Evidence-Based Guidelines for Depressive and Anxiety Disorders? *Med Care* 2010, **48**:240-8.
10. Houghton G, Rouse A: Are NHS primary care performance indicator scores acceptable as markers of general practitioner quality? *Br J Gen Pract* 2004, **54**:341-4.
11. Van Amstel RJ, Anema JR, Jettinghoff K, Verbeek JH, Nauta AP, Van Putten DJ: [Limited change in the quality of the social medical guidance and in the satisfaction of sick-listed patients, after collaborative projects between general practitioners and occupational physicians]. *Ned Tijdschr Geneeskde* 2005, **149**:2407-12.
12. Van der Weide WE, Verbeek JH, Van Dijk FJ. Relation between indicators for quality of occupational rehabilitation of employees with low back pain. *Occup Environ Med* 1999, **56**:488-93.
13. Advani A, Shahar Y, Musen MA. Medical quality assessment by scoring adherence to guideline intentions. *Proc AMIA Symp* 2001:2-6.

14. Carek PJ, Dickerson LM, Boggan H, Diaz V: A limited effect on performance indicators from resident-initiated chart audits and clinical guideline education. *Fam Med* 2009, **41**:249-54.
15. O'Malley AS, Clancy C, Thompson J, Korabathina R, Meyer GS: Clinical practice guidelines and performance indicators as related--but often misunderstood--tools. *Jt Comm J Qual Saf* 2004, **30**:163-71.
16. Rebergen D, Hoenen J, Heinemans A, Bruinvels D, Bakker A, Van Mechelen W: Adherence to mental health guidelines by Dutch occupational physicians. *Occup Med* 2006, **56**:461-8.
17. Smolders M, Laurant M, Verhaak P, Prins M, Van Marwijk H, Penninx B, et al.: Adherence to evidence-based guidelines for depression and anxiety disorders is associated with recording of the diagnosis. *Gen Hosp Psychiatry* 2009, **31**:460-9.
18. Wilson APR, Gibbons C, Reeves BC, Hodgson B, Liu M, Plummer D, et al.: Surgical wound infection as a performance indicator: agreement of common definitions of wound infection in 4773 patients. *BMJ* 2004, **329**:720.
19. Zwerver F, Schellart AJM, Anema JR, Rammeloo K, Van der Beek AJ: Intervention mapping for the development of a strategy to implement the insurance medicine guidelines for depression. *BMC Public Health* 2011, **11**:9.
20. Bijl RV, Ravelli A, Van Zessen G: Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Soc Psychiatry Psychiatr Epidemiol* 1998, **33**:587-95.
21. Levav I, Rutz W: The WHO World Health Report 2001 new understanding--new hope. *Isr J Psychiatry Relat Sci* 2002, **39**:50-6.
22. Ormel J, Petukhova M, Chatterji S, Guilar-Gaxiola S, Alonso J, Angermeyer MC, et al.: Disability and treatment of specific mental and physical disorders across the world. *Br J Psychiatry* 2008, **192**:368-75.
23. Wang F, Lee HP, Lu C. Thermal-mechanical study of functionally graded dental implants with the finite element method. *J Biomed Mater Res A* 2007, **80**:146-58.
24. Health Council: *Insurance medicine guidelines for depression*. Den Haag: Gezondheidsraad, 2006.
25. Shoukri MM, Asyali MH, Donner A: Sample size requirements for the design of reliability study: review and new results. *Stat Methods Med Res* 2004, **13**:251-271.
26. SPSS Inc: *SPSS 15.0 Command Syntax Reference*, 2006.
27. Molenberghs G, Laenen A, Vangeneugden T: Estimating reliability and generalizability from hierarchical biomedical data. *J Biopharm Stat* 2007, **17**:595-627.
28. Shavelson RJ, Webb NM, Rowley GL: Generalizability theory. *American Psychologist* 1989, **44**:922-32.
29. Shavelson RJ, Webb NM: *Generalizability theory: a primer*. Newbury, Park, London, New Delhi: Sage Publications, 1991.
30. Euser AM, Le Cessie S, Finken MJ, Wit JM, Dekker FW: Reliability studies can be designed more efficiently by using variance components estimates from different sources. *J Clin Epidemiol* 2007, **60**:1010-1014.
31. Hamilton M: A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960, **23**:56-62.
32. De Boer WEL, Wind H, Van Dijk FJH, Willems HHB: Interviews for the assessment of long-term incapacity for work: a study on adherence to protocols and principles. *BMC Public Health* 2009, **9**:169.
33. Van Rijssen HJ, Schellart AJM, Anema JR, Van der Beek AJ. A theoretical framework to describe communication processes during medical disability assessment interviews. *BMC Public Health* 2009, **9**:375.



34. Hepner KA, Rowe M, Rost K, Hickey SC, Sherbourne CD, Ford DE, et al.: The effect of adherence to practice guidelines on depression outcomes. *Ann Intern Med* 2007, **147**:320-9.
35. Verbeek J, Spelten E, Kammeijer M, Sprangers M: Return to work of cancer survivors: a prospective cohort study into the quality of rehabilitation by occupational physicians. *Occup Environ Med* 2003, **60**:352-7.
36. Hermanides HS, Hulscher MEJL, Schouten JA, Prins JM, Geerlings SE: Development of quality indicators for the antibiotic treatment of complicated urinary tract infections: a first step to measure and improve care. *Clin Infect Dis* 2008, **46**:703-11.
37. Spreeuwers D, De Boer AGE, Verbeek JHA, Van Beurden MM, Van Dijk FJH: Diagnosing and reporting of occupational diseases: a quality improvement study. *Occup Med* 2008, **58**:115-21.
38. Schellart AJM, Mulders HPG, Steenbeek R, Anema JR, Kroneman H, Besseling JMJ: Inter-doctor variations in the assessment of functional incapacities by insurance physicians. *BMC Public Health* 2011, **11**:864.
39. De Boer WEL, Besseling JMJ, Willems HHB: Organisation of disability evaluation in 15 countries. *Prat Organ Soins* 2007, **38**:205-17.

# **Chapter**

---

# **4**

## **Relationships between intention to use guidelines and determinants of insurance physician's behaviour**

**F Zwerver  
AJM Schellart  
JR Anema  
AJ van der Beek**

## Abstract

**Background:** We studied the intention of a group of insurance physicians to use the guidelines for depression. We considered attitude, social norm and self-efficacy, knowledge/skills and barriers/stimuli, based on the Attitude - Social norm - self-Efficacy model (ASE model) as possible determinants of that intention.

**Aims:** The aim of this study was to understand the determinants of insurance physicians' behaviour when they are expected to use guidelines in daily practice.

**Method:** A representative sample of 42 insurance physicians participated in this study. Cross-sectional data were collected by means of a questionnaire based on the ASE model. Analyses were performed with a structural equations model (LISREL).

**Results:** Important determinants of the intention, the self-reported use of the guidelines, and change in assessment behaviour were: the influence of colleagues, the self-efficacy of the insurance physicians, and the implementation of the guidelines. Intention to use the guidelines for depression was not associated with the self-reported use of these guidelines, but it was associated with self-reported change in assessment behaviour.

**Conclusions:** Almost all the insurance physicians in this study intended to use at least elements of the guidelines. Their guideline adherence was explored with help of the ASE model, showing associations between guideline adherence and various determinants, but the ASE model could only partly be confirmed.



## Introduction

Since the introduction of evidence-based guidelines in health care, the adherence of physicians to those guidelines has been subject of research [1-3]. The implementation of guidelines and the adoption of guidelines by physicians in daily practice appeared to be a complex process, much of which is still unknown [4]. Researchers have made various efforts, using different theories and methods, to explore the adherence of physicians to guidelines [4-6]. Researchers have often used the Theory of Planned Behaviour (TPB) and its derivative, the Attitude, Social norm, self-Efficacy model (ASE model) [7, 8], to investigate the behavioural aspects of the use of guidelines by physicians [9-13]. The aim of this present study is to explore and understand the physicians' behaviour towards guidelines with help of the ASE model.

The ASE model explains behaviour by linking attitude, social norm and self-efficacy with behavioural intention and actual behaviour [14]. In addition to the three determinants of behavioural intention and actual behaviour, interfering factors such as 'knowledge and skills' and 'barriers and stimuli' may play a role. The immediate precursor of behaviour is intention, but to predict whether a physician intends to use guidelines, we need to know the physician's attitude towards the guidelines. In the ASE model, intention is also determined by social-influence and self-efficacy. An individual physician might feel pressured by colleagues or a staff physician to use these guidelines. The degree of self-efficacy that a physician feels in applying guidelines can also determine his intention to use guidelines. According to the ASE model, the link between the intention and the actual use of the guidelines could be stronger if the use of the guidelines is promoted by facilitating factors, such as a multifaceted implementation strategy, or an easy accessibility of the guidelines. This link could be weaker if barriers occur between intention and use. For example, lack of practical applicability, lack of agreement, and lack of supporting staff are well-known barriers in adherence to guidelines [6]. The ASE model is presented in Figure 4.1.

Although researchers have succeeded in identifying barriers in the use of guidelines, and have recommended improvements in implementation strategies, the adherence of physicians to guidelines often remains low [1, 15-17]. Therefore, we need to gain more knowledge and insight into the entire process, from dissemination of the guidelines to use by physicians in daily practice.

The TPB and the ASE model have been used to explain, among other things, the behaviour of physicians and patients concerning guidelines in an occupational health context [12, 18, 19]. We are interested in the guideline adherence in the field of insurance medicine. Insurance physicians (IPs) working in the Netherlands, have recently been confronted with guidelines for the first time. An IP is a physician with a registration in insurance medicine, acquired after four years of post-graduate education. The IP assesses disability claims from employees by doing an interview, an examination, and

filling in a functional ability list, all recorded in a medical work disability report [20]. Insurance medicine guidelines cover the work disability assessment of an employee. Starting with the assessment of the diagnosis and co-morbidity, judgement of severity of the disorder, identification of risk factors, assessment of therapy, treatment, and participation efforts and ending with the assessment of the work limitations and the prognosis [21]. We chose to study the guidelines for depression, because world wide depression increasingly accounts for long-term disability [22, 23]. For the assessment of the diagnosis depression the guidelines refer to the Diagnostic Statistic Manual IV.

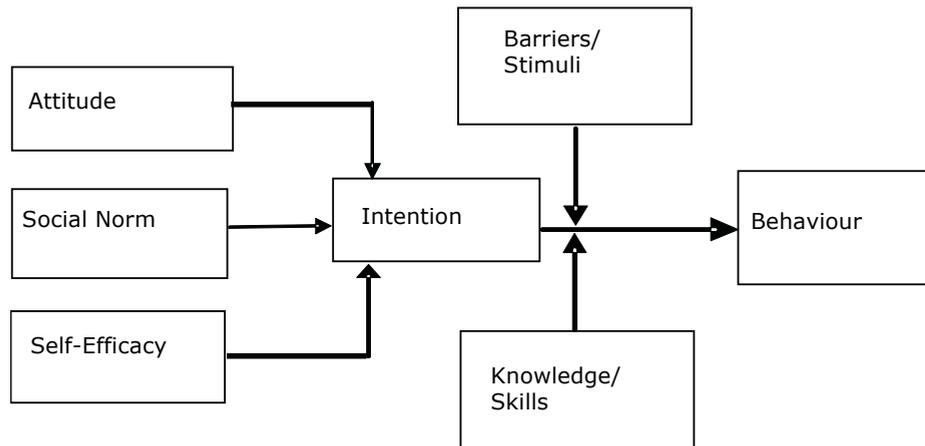


Figure 4.1: ASE-model [8].

We used the ASE model as a systemic framework for identification of behavioural antecedents to adherence to the guidelines, and the various stimuli or barriers that might influence the IPs' behaviour regarding to the guidelines. In this model, we hypothesized relations between intention and self-reported use of the guidelines, and subsequently between self-reported use of the guidelines and self-reported change of assessment behaviour due to applying the guidelines. The first research question was: what are the most important determinants of a) the intention to use the guidelines for depression, b) the self-reported use of these guidelines, and the c) self-reported change in their assessment behaviour? Secondly: is the intention to use the guidelines for depression associated with the self-reported use of these guidelines and the self-reported change in assessment behaviour? And finally: is there a relationship between the self-reported use of the guidelines for depression and the self-reported change in assessment behaviour?



Table 4.1: Constructs of the ASE model, and the background of the insurance physicians.

Constructs of ASE model	#
Attitude of the IPs to guidelines in general	9
Attitude of the IPs to the GD	9
Social influence of colleague IPs on the use of the GD	9
Social influence of important others (colleagues, supervisors and clients) on adherence to the GD	5
Self-efficacy of the IPs concerning understanding and use of the GD	11
Knowledge and skills of the IPs concerning the GD	8
Barriers/stimuli in the use of the GD concerning the structure and lay-out of the GD	3
Barriers/stimuli in use of the GD concerning the implementation of the GD at the Institute	3
Barriers/stimuli in the use of the GD concerning organizational factors within the Institute	9
Barriers/stimuli in the use of the GD concerning the tools provided by the Institute during the implementation of the GD	16
Barriers/stimuli in the use of the GD concerning the quality of the GD	11
Intention of the IPs to (more stringently) use, or pertaining to the use of the GD	10
Behaviour of the IPs concerning the use of the GD in daily practice	4
Changes in IP assessment behaviour influenced by the GD	3
Background of IPs	
Age (25-65 years, in 5-year classes)	1
Gender (male, female)	1
Years of experience as physician in general	1
Years of experience as IP	1
Hours per week working as IP	1
Registered as IP (no, yes)	1
Statutory background of the assessments of (the majority of) IP clients	5
Employed (versus temporarily hired) by the Institute	1
Industrial insurance boards (the predecessors of the Institute) the IPs had worked with before (7 possibilities)	1
Type of educational and informative activities to maintain the medical standards	6
Time needed for assessment of a client with depression (average)	1
Number of patients (with depression) assessed per month (average)	1

IP=Insurance physician; GD=Guidelines for Depression; Institute=the Dutch Institute for Employee Benefits Schemes; #=Number of items in the questionnaire.



## Method

We developed a questionnaire, with the ASE model as theoretical background [8], paying attention to the application of the guidelines for depression by IPs. The questions were based on, and in most cases adapted from research literature regarding guideline adherence, and focussed on the ASE model. The questionnaire included 14 theory based constructs from the ASE model, such as intentions, attitudes, social norm, self-efficacy, knowledge/skills, and barriers/stimuli in relation to guidelines in general and to the guidelines for depression in particular. Most of the questions were rated on a 5-point Likert scale, ranging from 'totally disagree' to 'totally agree'; others were rated on a 10-point rating scale. The questionnaire also contained questions about the background of the IPs. The constructs of the ASE model used in the questionnaire, the number of items for each construct, and the background of the IPs are summarized in Table 4.1. The questionnaire can be found in the Appendix.

IPs who worked at the Dutch Institute for Employee Benefits Schemes (Institute) were invited to follow a four-day postgraduate course in applying the guidelines in the disability assessment of clients with depression in the period from March to July 2009. The inclusion criteria were: registered as an IP or currently following the post-academic Insurance Medicine colloquium, and performing disability assessments of clients. 42 participating IPs filled in the questionnaire. The answers given by the 42 IPs were used to determine which constructs from the questionnaire were suitable for further analysis. Reliability analysis, including item-analysis, was performed for the 14 constructs of items that were theoretically assumed to form an additive scale. We considered a Cronbach's alpha of 0.65 as the minimum of internal consistency of a scale. These reliability analyses were performed in the SPSS 15.0 program. We recoded items on the questionnaire concerning the background of the IPs into 15 background variables. To select the possibly relevant background variables we used the Ordinary Least Squares -regression backward selection (Pin=0.05, Pout=0.10) option of the SPSS 15.0 program, with all background variables as independent variables and each of the 14 scale variables as dependent variable. Eight background variables that had a meaningful association with one or more scale variables were included in further analyses. The correlations between the 14 scale variables and the 8 background variables were calculated in Prelis 2.72 [24]. To interpret the relationships between the variables, we used Lisrel 8.72 [25] to examine the correlation matrix in a structural equations model with observed variables, i.e. a path model. Because the size of our study sample was too small to achieve a reliable model, we artificially increased the number of participants to  $N = 200$ , which can be considered as an optimum number [26]. By doing this, the estimated direct effects are not influenced, but the standard errors of these effects become smaller and the significance of the estimated effects becomes larger. Furthermore, the model fit becomes less and, assuming the same degrees of freedom, as a consequence the model has a greater chance to be rejected. Especially because of this last mentioned aspect we

think that an artificially increase of the number of participants to the optimum number is acceptable to explore relationships in a path model.

According to the theoretical ASE model, we formulated a structural start model with the scale variables as endogenous variables and the background variables as exogenous variables. Subsequently, with an estimated start model containing 11 endogenous scale variables and six exogenous background variables, we fitted the model, i.e. closing non-significant parameters between endogenous variables and opening parameters with significant modification indices ( $>3.84$ ) within the theory-based constraints. The model fit was good [27] if the (Normal Theory Weighted Least Squares) Chi-square of the model was small, i.e. less than twice the number of degrees of freedom (df), if the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR) were both less than 0.05, and if the Comparative Fit Index (CFI) was equal to or greater than 0.90. Furthermore, we verified that the Q-plot of the standardized residuals crossed the diagonal for normal distribution and that the correlation of estimates were not higher than 0.7. The Medical Ethical Committee of the VU University approved the study.

## Results

### Background variables

The self-reported background variables of the IPs are presented in Table 4.2. We checked whether the 42 IPs who participated in this study were a representative sample of the total group of IPs who worked for the Institute ( $n=900$ ) with respect to gender, age and working hours per week. In the group of participants the mean age was 51 years ( $SD = 14.90$ ) ( $CI = [46.5; 55.7]$ ), 47.8 % were female and they worked on average 31.68 hours per week ( $SD = 9.31$ ) ( $CI = [28.9; 34.5]$ ). The mean age of the total group was 49 years, 41.7% were female, and they worked on average 32 hours per week (distribution measures of the total group could not be calculated) [28]. The mean age and the number of hours worked by an IP of the total group were within the 95% CI of the participants' group, which appeared to be a representative sample.

They assessed on average, 7 clients with depression ( $SD 5.0$ ) per month. Almost all of them reported that they intended to use, or continued to use certain elements of the guidelines for depression. Approximately 50% reported that they used the complete guidelines for depression, and approximately 85% reported that they used at least some elements of the guidelines for depression. Approximately 50% of the IPs reported that they had more or less changed their assessment behaviour because of these guidelines.

**Table 4.2:** Background variables of insurance physicians (n=42).

Description of the background variables	%
<b>Age in years<sup>§*</sup></b> (mean = 51.10; sd = 6.34)	
50 years and younger = 1	45.2
Older than 50 years = 2	54.8
<b>Number of working hours per week<sup>§</sup></b> (mean = 31.69; sd = 9.31)	
Part-time ( $\leq$ 34 hours) = 1	47.6
Full-time ( $>$ 34 hours) = 2	52.4
<b>Number of clients with depression per month<sup>§</sup></b> (mean = 7.02 clients; sd = 4.98)	
0 - 1 client = 1	9.5
2 - 4 clients = 2	19.0
5 - 8 clients = 3	38.1
9 - 10 clients = 4	21.4
11 or more clients = 5	11.9
<b>Assessment time for depressed clients<sup>§*</sup></b> (mean = 144.5 minutes; sd = 54.41)	
30 - 45 minutes = 1	4.8
46 - 110 minutes = 2	19.0
111 - 180 minutes = 3	57.1
181 - 210 minutes = 4	14.3
211 - 240 minutes = 5	4.8
<b>Years working as physician<sup>§*</sup></b> (mean = 22.67 years; sd = 5.65)	
10 - 18 years = 1	23.8
19 - 27 years = 2	52.4
28 - 31 year = 3	23.8
<b>Years working as insurance physician<sup>§*</sup></b> (mean = 15.38 years; sd = 7.79)	
6 - 9 years = 1	26.2
10 - 22 years = 2	50.0
23 - 31 years = 3	23.8
<b>Intensity of kind of professional activities<sup>§*</sup></b> (mean = 3.95 activities; sd = 0.79)	
1-3 activities = 1	23.8
4 activities = 2	52.4
5 activities = 3	23.8
<b>Gender</b>	
Male = 1	52.4
Female = 2	47.8
<b>Registered as insurance physician*</b>	
Yes = 1	85.7
No = 2	14.3
<b>Employee of the Institute*</b>	
Yes = 1	78.6
No = 2	21.4

§Recoded into an ordinal variable; \*Not used in the structural equations model; Institute=Dutch Institute for Employee Benefits Schemes.

### ASE scale variables

The 14 ASE scale variables for the 42 IPs are presented in Table 4.3. The average Cronbach's alpha of the scales was 0.79.

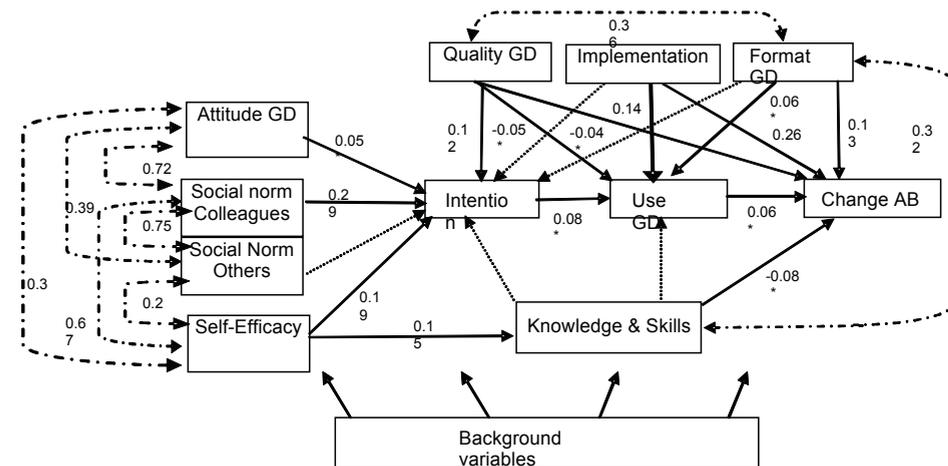
**Table 4.3:** ASE scale variables.

ASE scale variables	Description	#	Theoretical (empirical)					
			Min	Max	Median	Mean	SD	
Attitude*	Attitude to guidelines in general	9	9 (18)	45 (44)	32.00	31.50	6.09	0.76
Attitude GD	Attitude to the GD	9	9 (19)	45 (44)	33.00	33.14	5.33	0.77
Social Norm Colleagues	Influence of colleagues on acceptance of the GD	9	9 (15)	29 (29)	22.50	22.74	3.67	0.69
Social Norm Others	Influence of important others in adherence to the GD	5	5 (5)	25 (22)	16.00	15.38	4.36	0.81
Self-Efficacy	Self-efficacy in using the GD	11	11 (24)	55 (48)	36.00	35.31	5.38	0.75
Knowledge & Skills	Knowledge of and skills with regard to the GD	8	8 (16)	40 (39)	28.00	27.81	5.34	0.77
Format GD	Stimulus in the use of the GD by the format of the guideline	3	3 (13)	30 (27)	20.00	19.98	3.11	0.90
Implementation <sup>§</sup>	Stimulus in the use of the GD by the implementation of the guideline	3	1 (1)	5 (5)	3.00	3.10	1.14	0.71
Institute*	Stimulus in the use of the GD by organisational factors within the Institute	9	9 (10)	45 (38)	26.00	25.38	6.97	0.84
Tools <sup>§*</sup>	Stimulus in the use of the GD by the tools delivered with the guideline	16	1 (1)	5 (5)	3.00	2.98	1.14	0.89
Quality GD	Stimulus in the use of the GD by the quality of the guideline	11	11 (24)	55 (51)	40.00	39.02	6.10	0.84
Intention	Intention to use the GD	10	10 (19)	50 (46)	35.00	35.05	5.80	0.76
Use GD <sup>§</sup>	Use of the GD	4	1 (1)	5 (5)	3.00	3.02	1.16	0.65
Change AB <sup>§</sup>	Change in assessment behaviour due to the GD	3	1 (1)	3 (3)	2.00	1.98	0.60	0.86

Scores of insurance physicians (n=42). #=number of items in the questionnaire; Min=minimum of the scales; Max=maximum of the scales; median of the scores; SD=standard deviations; =reliability of the scales (Cronbach's alpha); GD=Guidelines for Depression; Institute=the Dutch Institute for Employee Benefits Schemes; §Recoded into an ordinal variable; \*Not used in the structural equations model.

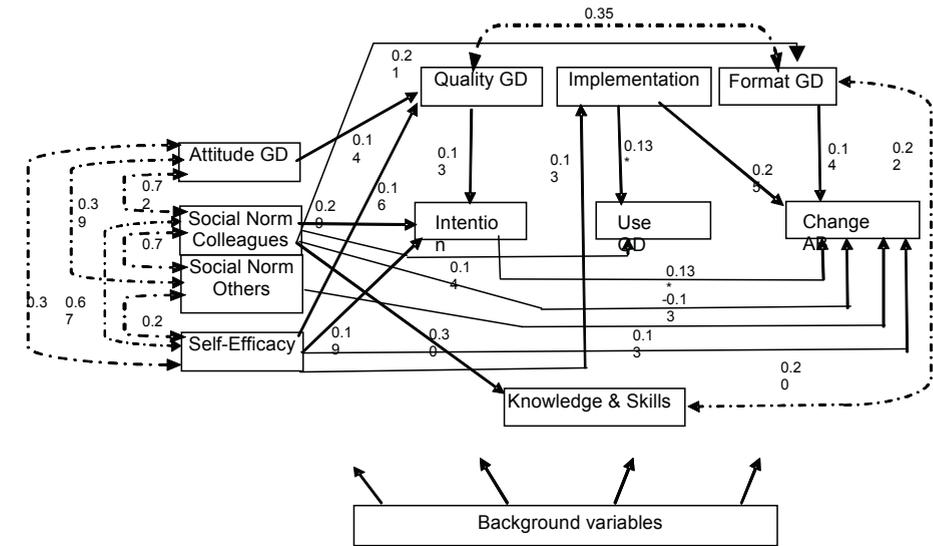
### Structural equations

The relationships in the start model between the endogenous variables (beta and psi matrix) are presented in Figure 4.2, together with the model fit parameters. These parameters indicated that the start model did not fit well (SRMR = 0.067; modification indices > 3.84). Direct effects of six background variables on endogenous variables in the model (not shown here) were significant: gender, number of clients with depression assessed by an IP per month, number of working hours per week by an IP, IPs' employment history and two different types of clients' disability benefit legislation.



**Figure 4.2:** Start model. Model fit parameters: Chi-square 91.726,  $df=73$ ,  $p = 0.068$ ; RMSEA = 0.036; SRMR = 0.067, CFI = 0.974. Straight lines indicate direct effects (beta matrix), double-headed dotted lines indicate associations in the disturbance terms (psi matrix); single arrowed dotted lines indicate direct effects which had to be included in the model to get positive definite matrices. All coefficients  $p < 0.05$ ; except (\*),  $p$  is not significant, at  $N=200$ , artificially. GD=Guideline Depression; AB=Assessment Behaviour.

We adapted the direct effect between the endogenous variables in the model to obtain good fit parameters. The final model, which contained the same direct effects of exogenous variables on the endogenous variables as in the start model, is presented in Figure 4.3, and shows the direct effects between the endogenous variables (beta matrix) and the associations between the (disturbance terms of) endogenous variables (psi matrix). The model fit parameters, and other parameters (Q-plot, modification indices, correlation of estimates), indicated that the model fit was good. The explained variance was highest for intention to use the guidelines (0.25) and for self-reported change in assessment behaviour (0.30).



**Figure 4.3:** Final Model. Model fit parameters: Chi Square 45.873,  $df=72$ ,  $p = 0.994$ ; RMSEA = 0.0; SRMR = 0.0411, CFI = 1.000. All coefficients  $p < 0.05$ ; except (\*),  $p < 0.10$ , at  $N=200$ , artificially. GD=Guideline Depression; AB=Assessment Behaviour. Explained variance ( $R^2$ ) of endogenous variables: Attitude to the GD (0.10); Social Norm, influence of colleagues on acceptance of the GD (0.05); Social Norm, influence of important others in adherence to the GD (0.10); Self-efficacy in the use of the GD (0.14); Knowledge and skills concerning the GD (0.16); Stimulus in the use of the GD due to the format of the guideline (0.10); Stimulus in the use of the GD due to the implementation of the guidelines (0.16); Stimulus in the use of the GD by the quality of the guideline (0.14); Intention to use the GD (0.25); Use of the GD (0.14); Change in assessment behaviour due to the GD (0.30).

With this final model, presented in Figure 4.3, we were able to answer the research questions. The answer to the first research question is: the important determinants with regard to a) *intention* to use the guidelines were: the influence of colleagues on acceptance of the guidelines, self-efficacy, and to a lesser extent the perceived quality of the guidelines. The important determinants with regard to b) *self-reported use* of the guidelines were: the influence of colleagues on acceptance of the guidelines, and to a lesser extent the stimulus in using the guidelines due to the implementation strategy. Important determinants with regard to c) *self-reported change* in assessment behaviour were: self-efficacy in using the guidelines, stimulus in using the guidelines due to the implementation strategy, and to a lesser extent the influence of colleagues and important others on acceptance of the guidelines. *Self-reported change* was also weakly positively stimulated by the lay-out of the guidelines.

The answer to the second and third research questions are: the final model showed no direct relation between *intention* and the *self-reported use* of the guidelines. Only a weak relationship between *intention to use* and *change in assessment behaviour* was found. We also found no relationship between *change in assessment behaviour* and the *self-reported use* of the guidelines. Instead, we found associations among the ASE determinants themselves.



## Discussion

Almost all of the participating IPs reported that they intended to use certain elements of the guidelines for depression. In addition, approximately 50% of the IPs changed their assessment behaviour due to the guidelines. The influence of colleagues on acceptance of the guidelines, self-efficacy in using the guidelines, and stimulus in using the guidelines due to their implementation, appeared to be the most important determinants of the intention to use the guidelines. Having found no association between change in assessment behaviour and use of the guidelines, although this was theoretically expected, can be explained by the fact that the IPs in our study possibly assumed that they already worked in accordance with the guidelines, and therefore they did not feel any need to change their assessment behaviour. At the Institute IPs are managed by their Staff IP in small groups, and not to our surprise we found that the IPs were influenced by the opinions of their colleagues or staff IPs. Self-efficacy is a personal determinant for intention to use the guidelines, which could be improved by further education. We also found that the implementation strategy for the guidelines influences guideline adherence. This provides opportunity for improvement in guideline adherence, by empowerment of the implementation strategy.

Strength of this study was that we used a theoretical psychological model to describe the behaviour of the IPs with regard to the use of guidelines. The questionnaire we developed included all relevant constructs from the ASE model. Furthermore, the scales we formulated for these constructs had moderate to good reliability, which made it possible to analyze the relationships between the ASE constructs as observed variables in a structural model.

The main limitation of this study is the cross-sectional design. Despite the use of Lisrel, which analyses associations of variables to determine cause and effect, we could draw no causal conclusions. The results of this study have to be interpreted with caution because of the self-reported nature of the study, the small number of participants, and the fact that we had to artificially increase the sample. Furthermore, the IPs participated voluntarily in the course for learning to apply the guidelines, which means that selection bias is possible.

The self-reported adherence to guidelines among the IPs appeared to be high in our study, compared to reports about other disciplines, such as primary care, clinical care, and occupational health care. We found an adherence of 85% to the guidelines, but this high percentage may (partly) be due to the legal obligatory character of the guidelines in general for insurance medicine. Former research among IPs in the Netherlands reported an adherence of 90% to the protocols for semi-structured assessment interviews in disability assessments [29]. In primary care, the overall adherence to 70 guidelines within a period of 10 years was 67% [16], whereas another study reported a low adherence (39%) to the guidelines for mental health problems by occupational physicians [12].

In general, it was suggested that 30-40% of the patients do not receive care according to current scientific evidence [30]. A review of 30 studies focussing on the attitude of clinicians towards guidelines, reported high satisfaction with clinical practice guidelines, but there were concerns about the applicability of the guidelines, their role in cost-reduction, and their potential for increasing litigation [31]. To our knowledge, there is only one other study that has focussed on the attitude of IPs towards guidelines [32]. However, this study was carried out in Belgium, where there are no specific insurance medicine guidelines. The conclusion in that study was that the attitude of the IPs towards clinical guidelines was positive, but their use of the guidelines was low. Studies, in which psychological theoretical models, such as the TPB or the ASE model, were used to describe the behaviour of physicians with regard to guidelines, resulted mostly in the conclusion that the model that was used only accounted for a small part of the variance in the adherence of the physicians to these guidelines [11, 17]. In our study we could only confirm this conclusion. However, as in other studies, the theoretical psychological model did appear to be helpful in describing the behaviour of the IPs with regard to the use of the guidelines [5, 11]. We now have indications as to what forces stimulate the use of the guidelines by IPs, and what barriers might occur. Unlike other studies, in which the behaviour of the physicians was described with the help of a theoretical psychological model [5, 10-12, 18, 19], we used structural equation modelling, which provided more insight into the complex processes of the behaviour of IPs, and changes in that behaviour when they are expected to work according to guidelines in daily practice. The paths formed by the associations that we found between the variables and the determinants of the ASE model showed us how IP adherence to the guidelines can be influenced, and how it can be improved.

Because the physicians in the present study are influenced by colleagues in their use of the guidelines, it makes sense to monitor their behaviour with regard to guideline adherence with specific performance indicators, and to provide them with feedback. Although the self-reported (intention to) use the guidelines in our study was already high, improvement in adherence to the guidelines could be achieved by increasing the self-efficacy of the IPs. We found that IPs with more self-efficacy were more receptive to changing their behaviour in order to apply guidelines in daily practice. In addition, given the positive association that we found between implementation and self-reported use of the guidelines and self-reported change in assessment behaviour, increased efforts to improve the implementation might result in an increase in IPs' guideline adherence. That may be a good starting point for interventions aimed at increasing guideline adherence. The most important ASE determinants of intention to use the guidelines for depression, the self-reported use of the guidelines by IPs, and change in their assessment behaviour, seem to be influenced by colleagues, the self-efficacy of the IPs, and various barriers/stimuli occurring in the implementation of the guidelines.



## Conclusions

Guideline adherence of insurance physicians was explored with help of the ASE model, showing associations between guideline adherence and various determinants, but the ASE model could only partly be confirmed. Important determinants for the intention to use guidelines for insurance physicians were: the influence of colleagues, self-efficacy and the implementation of the guidelines. Intention to use the guidelines was associated with change in assessment behaviour, and 50% of the insurance physicians had changed their assessment behaviour due to the implementation of the guidelines for depression. We see opportunities to improve insurance physicians' guideline adherence by offering them a multifaceted training in which they learn to apply the guidelines for depression.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

The authors declare that they participated in the study and made the following contributions to the study, and that they have seen and approved the final version. FZ and AJMS wrote the manuscript. AJMS and FZ performed the analysis. AJMS, FZ, JRA, and AJvdB contributed to the conception and design of this study. JRA and AJvdB revised and commented on the manuscript. AJMS, JRA and AJvdB will act as guarantors of this study.

### Acknowledgements and Funding

The authors wish to thank the insurance physicians who participated in this research. The authors like to thank K.M. van Beurden Msc for her assistance in the development of the questionnaire used in his study. The Research Center for Insurance Medicine AMC-UMCG-UWV-VU University Medical Center, in Amsterdam, is a joint initiative of the Academic Medical Center (AMC), the University Medical Center in Groningen (UMCG), the Dutch Institute for Employee Benefits Schemes (UWV), and the VU University Medical Center (VUMC). This work was supported by the Dutch Institute for Employee Benefits Schemes. [Grant number 0Z06094002]. The study sponsor had no (decisive) role in the study design, in the data-collection, data-analysis, or interpretation of the data, in the writing of the report, or in the decision to submit the paper for publication.

## References

1. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al.: Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 1999, **282**:1458-65.
2. Davis DA, Taylor-Vaisey A: Translating guidelines into practice. A systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines. *CMAJ* 1997, **157**:408-16.
3. Lomas J, Anderson GM, Domnick-Pierre K, Vayda E, Enkin MW, Hannah WJ: Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. *N Engl J Med* 1989, **321**:1306-11.
4. Grol R, Wensing M: What drives change? Barriers to and incentives for achieving evidence-based practice. *Med J Aust* 2004, **180**:S57-S60.
5. Bonetti D, Johnston M, Pitts NB, Deery C, Ricketts I, Tilley C, et al.: Knowledge may not be the best target for strategies to influence evidence-based practice: using psychological models to understand RCT effects. *Int J Behav Med* 2009; **16**:287-93.
6. Cabana MD, Rushton JL, Rush AJ: Implementing practice guidelines for depression: applying a new framework to an old problem. *Gen Hosp Psychiatry* 2002, **24**:35-42.
7. Ajzen I: The theory of planned behavior. *Organ Behavi and Hum Decis Process* 1991, **50**:179-211.
8. De Vries H, Dijkstra M, Kuhlman P: Self-efficacy: third factor besides attitude and subjective norm as a predictor of behavioral intention. *Health Educ Res* 1988, **3**:273-82.
9. Brouwer S, Krol B, Reneman MF, Bultmann U, Franche RL, Van der Klink JIL, et al.: Behavioral determinants as predictors of return to work after long-term sickness absence: an application of the theory of planned behavior. *J Occup Rehabil* 2009, **19**:166-74.
10. Ceccato NE, Ferris LE, Manuel D, Grimshaw JM: Adopting health behavior change theory throughout the clinical practice guideline process. *J Contin Educ Health Prof* 2007, **27**:201-7.
11. Hanbury A, Wallace L, Clark M: Use of a time series design to test effectiveness of a theory-based intervention targeting adherence of health professionals to a clinical guideline. *Br J Health Psychol* 2009, **14**:505-18.
12. Rebergen D, Hoenen J, Heinemans A, Bruinvels D, Bakker A, Van Mechelen W: Adherence to mental health guidelines by Dutch occupational physicians. *Occup Med* 2006, **56**:461-8.
13. Van Oostrom SH, Anema JR, Terluin B, Venema A, De Vet HCW, Van Mechelen W: Development of a workplace intervention for sick-listed employees with stress-related mental disorders: Intervention Mapping as a useful tool. *BMC Health Serv Res* 2007, **7**:127.
14. De Vries H, Backbier E, Kok G, Dijkstra M: Measuring the impact of social influences on smoking onset in a longitudinal study: an integration of social psychological approaches. *J Appl Soc Psychol* 1995, **25**:237-57.
15. Grol R, Wensing M, Eccles MP: *Improving Patient Care: the implementation of change in clinical practice*. Oxford: Elsevier, 2005.
16. Grol R: Successes and failures in the implementation of evidence-based guidelines for clinical practice. *Med Care* 2001, **39**:1146-1154.
17. Maue SK, Segal R, Kimberlin CL, Lipowski EE: Predicting physician guideline compliance: an assessment of motivators and perceived barriers. *Am J Manag Care* 2004, **10**:383-91.
18. Van Oostrom SH, Anema JR, Terluin B, De Vet HCW, Knol DL, Van Mechelen W: Cost-effectiveness of a workplace intervention for sick-listed employees with common mental disorders: design of a randomized controlled trial. *BMC Public Health* 2008, **8**:12.



19. Vermeulen S, Anema J, Schellart A, Van Mechelen W, Van der Beek A: Intervention mapping for development of a participatory return-to-work intervention for temporary agency workers and unemployed workers sick-listed due to musculoskeletal disorders. *BMC Public Health* 2009, **9**:216.
20. Spanjer J, Krol B, Popping R, Groothoff JW, Brouwer S: Disability assessment interview: the role of detailed information on functioning in addition to medical history-taking. *Disabil and Rehab* 2008, **30**:1885-90.
21. De Boer WEL, 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 developing to AGREE in Germany and the Netherlands. *BMC Public Health* 2009, **9**:349.
22. Ormel J, Petukhova M, Chatterji S, Auilar-Gaxiola S, Alonso J, Angermeyer MC, et al.: Disability and treatment of specific mental and physical disorders across the world. *Br J Psychiatry* 2008, **192**:368-375.
23. Zwerfer F, Schellart AJM, Anema JR, Rammeloo KR, Van der Beek AJ: Intervention mapping for the development of a strategy to implement the insurance medicine guidelines for depression. *BMC Public Health* 2011, **11**: 9.
24. Jöreskog KG, Sörbom D: *Prelis™ 2: User's Reference Guide*. Lincolnwood, Illinois SSI Inc, 2002 (3<sup>e</sup> edition).
25. Jöreskog KG, Sörbom D: *LISREL® 8: User's Reference Guide*. Chicago, Illinois, SSI Inc, 2001.
26. Hayduk LA: *Structural equation modelling with Lisrel: essentials and advances*. Baltimore: The Johns Hopkins University Press, 1987.
27. Hooper D, Coughlan J, Mullen M R: Structural equation modelling: guidelines for determining model fit. *EJBRM* 2008, **6**:53 – 60.
28. Van Rijssen HJ, Schellart AJ, Anema JR, Van Der Beek AJ: Determinants of physicians' communication behaviour in disability assessments. *Disabil Rehabil* 2011, **33**:1157-68.
29. De Boer WEL, Wind H, Van Dijk FJH, Willems HHB: Interviews for the assessment of long-term incapacity for work: a study on adherence to protocols and principles. *BMC Public Health* 2009, **9**:169.
30. Grol R, Grimshaw J: From best evidence to best practice: effective implementation of change in patients' care. *Lancet* 2003, **362**:1225-30.
31. Farquhar CM, Kofa EW, Slutsky JR: Clinicians' attitudes to clinical practice guidelines: a systematic review. *Med J Aust* 2002, **177**:502-6.
32. Heselmans A, Donceel P, Aertgeerts B, van de Velde S, Ramaekers D: The attitude of Belgian social insurance physicians towards evidence-based practice and clinical practice guidelines. *BMC Fam Pract* 2009, **10**:64.

# Chapter

---

# 5

## **An implementation strategy for insurance physicians to improve guideline adherence: an experiment in a controlled setting**

*Implement Sci* 2011, 6:131

**F Zwerver  
AJM Schellart  
DL Knol  
JR Anema  
AJ van der Beek**

## Abstract

**Background:** The aim of this study was to investigate the efficacy of a newly developed implementation strategy for the insurance medicine guidelines for depression in the Netherlands. We hypothesized that an educational intervention would increase the insurance physicians' guideline adherence in a controlled setting.

**Methods:** Forty Insurance physicians were allocated in a randomised controlled trial (RCT) to an intervention group (n = 21) and a control group (n = 19). The intervention group received tailored training in applying the guidelines for depression, while the control group received an alternative programme. Baseline (T0) and follow-up (T1) measurements were conducted before and after the intervention within a period of two weeks. The intervention consisted of a workshop in which the evidence-based theory of the guidelines was translated for use in practice, with the help of various tools. The insurance physicians had to write a case-report on the basis of video cases, two before and two after the training. Specially trained and blinded Test insurance physicians judged the case-reports independently on the basis of six performance indicators. Primary outcome measure in the controlled setting of the trial was guideline adherence measured by six performance indicators on a scale of one to seven. Secondary outcome measure was knowledge of the guidelines for depression. Analyses were performed using Linear Mixed Models, and ANCOVA.

**Results:** We found significantly higher scores in the intervention group than in the control group at T1 for both outcomes. The interaction effect (standard error; p-value) of group crossed with time was 0.97 (0.19;  $p < 0.0005$ ) for guideline adherence in the controlled setting. The group effect at T1 for the knowledge test was 0.86 (0.40;  $p = 0.038$ ).

**Conclusions:** The newly developed implementation strategy for the insurance medicine guidelines for depression improved the guideline adherence of the trained insurance physicians in disability assessments of clients with depression when performed in a controlled setting. Furthermore, the trained insurance physicians showed gains in knowledge of the guidelines for depression.



## Background

The implementation of guidelines, together with the difficulties and barriers that might occur when evidence-based medicine has to be translated into a physician's daily practice, has been the subject of various studies [1-4]. Results have suggested a sizeable gap between the ideal and the actual performance of physicians in the application of guidelines [5] and explanations for this gap have emerged. In existing studies barriers to effective guideline implementation have been analysed at the level of the patient, the physician or the health-care organization [1, 3, 4]. In this study we put the focus on the physician's guideline adherence. Our aim is to investigate in a controlled setting the influence of a newly developed implementation strategy on the adherence of the insurance physician (IP) to the depression guidelines. IPs evaluate the disability claims of employees, which is of great societal and financial importance; they write their assessments in a medical work disability report and the benefit level of the employee is defined on the basis of that report. We chose to study the insurance medicine guidelines for depression because depression accounts for a substantial and increasing proportion of long-term disability claims; this corresponds with worldwide trends which show that depression is expected to be one of the leading causes of disability-adjusted life years in 2020 [6, 7].

We questioned whether the usual implementation of the guidelines for depression could be improved. After research, in which we used the Intervention Mapping method we developed a newly implementation strategy for these guidelines [8]. It was found that this implementation strategy should include an interactive educational training for IPs and tools, with the aim to learn them and facilitate them to apply these guidelines [8].

This aim is challenging because it means overcoming potential barriers such as negative aspects of physicians' behaviour in the adoption of guidelines. One of the summary points mentioned in an overview of reviews concerning the gap between research and practice is that passive dissemination of guidelines is generally ineffective [9, 10]. Nevertheless, change in the behaviour of physicians was found for certain educational interventions, while in other more formal types of medical education there was no evidence of change [10]; interactive sessions that enhance participant activity and provide the opportunity to practice skills for instance were found to result in changes in the physician's performance [11]. In the development of our strategy, we took these findings into consideration by consulting educational experts and by assessing the needs of the IPs [8]. The IPs wanted tools such as a checklist or a guideline summary card to facilitate the use of guidelines in practice. The experts expected a multifaceted strategy with interactive educational sessions and the practice of skills to be most effective. After taking these findings into account we developed an implementation strategy and evaluated its efficacy by setting up an RCT in which we compared implementation of the

guidelines using this strategy with the usual levels of implementation in the Netherlands [12, 13]. Primary outcome measure was the guideline adherence of the IP in a controlled setting, leading to the research questions: does training IPs in applying the guidelines for depression improve their guideline adherence in the work disability assessment of clients with depression in a controlled setting? Additionally, does the strategy improve their knowledge of the guidelines for depression?

The Medical Ethics Committee of VU University Medical Center approved the study design and the Netherlands Trial Registration accepted the RCT: NTR1863.

## Methods

### Design

To determine the efficacy of training IPs in applying the guidelines for depression, we conducted an RCT in which we compared the intervention implementation strategy with the usual methods of training IPs, by measuring their performances in disability assessments of clients with depression. The intervention was a training programme designed for IPs, in which they learned to apply the guidelines for depression. This programme, together with the baseline and follow-up measurements, was integrated into a four-day post-graduate course located at the Netherlands School of Public and Occupational Health (NSPOH). At the NSPOH we created a controlled setting in which we carried out the RCT. While the intervention group (IG) was trained in applying the guidelines for depression, the control group (CG) received an alternative programme of training in motivational interviewing, which did not conflict with the intervention programme. The RCT took three days within a period of two weeks in March 2009. After the RCT had been ended, for reasons of recruitment and equal treatment for both groups, the control group received the same training as the intervention group, while the intervention group received the alternative programme; this was planned as the fourth day of the course which was held three months later at the end of June 2009. At the NSPOH we managed to create a laboratorial setting where we could measure each IP's work disability assessments of clients with depression played by actors on video. The training programme was developed for use in practice.

### Study population and recruitment procedure

In January 2009, IPs working at the Dutch National Institute for Employee Benefits (the Institute) were invited to attend a post-graduate course in applying disability assessments of clients with depression in the period from March to July 2009. The Institute is responsible for evaluating disability claims of employees. The inclusion criteria were being registered as an IP or still following the post-academic colloquium on insurance medicine and conducting disability assessments of clients as commissioned



by the Institute. Participation was voluntary. The NSPOH was responsible for enrolling participants in the course during the period January to March 2009.

### Allocation and blinding

The IPs who participated in the post-graduate education programme were individually allocated in order of registration to the intervention group or to the usual care group by means of a random-sequence table. To prevent an unequal allocation across both groups, the participants were stratified before randomisation on three prognostic factors: age, gender and registration as an IP. The randomisation and stratification procedure was executed by a research assistant. After the stratification and randomisation procedure the dates for the RCT were planned in cooperation with the NSPOH. Participants were assigned to either the intervention or the control group by the research assistant, while the assignment was communicated to the participants by the NSPOH. Participants who were not available on the planned dates were excluded from the trial. The participants were blinded for the complete trial, including baseline, intervention or usual implementation programme, and follow-up. The participants were informed about the fact that the course was part of a research project, but they were not informed about the design and were blinded for the type of group they participated in. Contamination between groups was not possible due to the design of the trial. The researchers were blinded for the collection of data. The data were coded by an independent research assistant.

### Measurements and data collection

Data were collected at the location of the NSPOH where the course took place. At baseline and at follow-up each IP assessed the disability of two clients, played by actors, who were presented separately on video. The actors played clients with depression which were reconstructed after real cases. The actors played their roles on the basis of extensive scripts, with room left for improvisation. Thereby, the actors realistically represented actual clients with depression. The videos showed the disability assessment encounter of a client and an independent IP (not a participant in the RCT), who was briefed to perform the assessment in complete accordance with the guidelines for depression. The decision phase of the assessment was not shown on the video. The IPs wrote their medical work disability reports immediately after watching each client on the video. All the medical work disability reports were collected during the RCT. Furthermore, each IP's knowledge of the guidelines for depression was measured at the start and at the end of the intervention and the control programme with the same set of knowledge questions.

## Outcomes

The primary outcome was the guideline adherence of IPs in the four work disability reports of clients with depression in the controlled setting of the RCT. We used performance indicators (PIs) to measure guideline adherence. The guideline adherence in these medical work disability reports is a proxy for the quality of the work disability assessments carried out by the IPs, given the guidelines for depression. The guideline adherence was measured by the six PI scale scores (range: 1-7). The six PIs were made up in the form of different decision trees with logic branches coming to an end with a score of NA or A. The main elements of the guidelines for depression, which should be detectable in an IP's work disability report on a client with depression, were covered by these PIs. A detailed description of the development and the reliability of the PIs can be found elsewhere [14]. In Table 5.1 the subjects of the six PIs are summarized. Furthermore, as a form of sensitivity analysis, we analyzed the guideline adherence in a different way as well i.e. by using PIs scored as a binary outcome, not adequate (NA) or adequate (A).

**Table 5.1:** Subjects of performance indicators for the guidelines for depression.

PI 1	Correct diagnosis <ul style="list-style-type: none"> <li>• DSM-IV criteria for depressive disorder</li> </ul>
PI 2	Determination of severity of the disorder <ul style="list-style-type: none"> <li>• Source: medical examination or e.g., information of curative physician, HRSD</li> <li>• Relation between severity of the disorder and the limitations</li> </ul>
PI 3	Origin, course and prognosis of the disorder <ul style="list-style-type: none"> <li>• Risk factors for depressive disorder</li> <li>• Course of depressive disorder</li> <li>• Substantiated prognosis of depressive disorder</li> </ul>
PI 4	Co-morbidity <ul style="list-style-type: none"> <li>• Presence or absence of co-morbidity</li> <li>• Influence of co-morbidity on prognosis and limitations</li> </ul>
PI 5	Evaluation of care and cure <ul style="list-style-type: none"> <li>• Level of information about claimant and medical treatment</li> <li>• Action for required information if necessary</li> <li>• Reasons for stagnation in recovery of functioning</li> <li>• Medical treatment related to rehabilitation</li> </ul>
PI 6	Assessment of work limitations <ul style="list-style-type: none"> <li>• Work limitations related to the severity of depressive disorder</li> <li>• Work limitations substantiated to insurance medicine standards</li> </ul>

Adapted from Schellart et al. 2011 [14].



The secondary outcome was the IP's knowledge of the guidelines for depression, which was measured with a knowledge test. The knowledge test was developed on the basis of the guidelines for depression. This test was piloted by two independent IPs, who were researchers from other research groups and not involved in our project. In the final version, the test consisted of 10 propositions derived from the guidelines to be scored true or false. Sum scores were calculated immediately before and after the intervention or control programme.

## Judgement of the medical work disability reports with PIs by Test IPs

After the RCT was completed, the medical work disability reports of all participants were measured with PIs at baseline and at follow-up by three pairs of senior Test IPs. These Test IPs had received separate training in which they learned to measure medical work disability reports using the PIs. Each medical work disability report from each participant from the RCT was judged blind by one pair of Test IPs, which produced the PI scores per report. In cases of disagreement in the pair of Test IPs about a certain PI score, a consensus procedure was followed, resulting in one PI score. We measured the guideline adherence by the scoring of the six PIs as a scale. Taking into account the distance between the NA scores and A scores within the PIs, and the differences across the PIs, the scores were recoded in the following way to form the scale: (NA1=1), (NA2=2), (NA3=3), (NA4..NA7=4), (A1=5), (A2=6), (A3, A4=7). A more extensive explanation of the scale and its reliability with two Test IPs is published elsewhere [14].

## Intervention implementation strategy

The implementation strategy was developed on the basis of a needs assessment carried out by IPs and of semi-structured interviews held with psychiatrists, researchers, IP trainers with experience in post graduate education and the psychiatrist who was member of the board that drew up the guidelines for depression. In this needs assessment the needs of the IPs concerning the implementation of the guidelines were investigated. Their needs and the recommendations of the other experts were integrated into the intervention strategy using Intervention Mapping [8]. This intervention strategy consisted of a specific training programme for IPs in which they learn to apply the guidelines for assessing depression. Several evidence-based tools were developed to serve this goal intended to improve the applicability of the guidelines. Realistic cases of clients with depression presented at a video screening were used to enrich the training. Learning objectives for the IPs were: to use the tools for the improvement of their diagnostic skills, to improve their performances in the work disability assessment of clients with depression, and to write their findings and conclusions down in well-argued medical reports. The aim was to bring the IPs' reports more in concordance with the guidelines, i.e. transparent, more evidence-based and their reports should contain well-argued conclusions of clients' limitations in working ability. The participant's self-activation was stimulated in an

interactive learning process with feedback given by the trainers. Two of the authors (FZ and JRA), both IPs themselves, were involved as trainers in the intervention programme. Appendix 5.1 gives an overview of the intervention programme.

### Sample size and measures

Sample size estimation was based on the minimum desirable change in the primary outcome that is the least sensitive, i.e. the NA/A-score of the guideline adherence. For detecting a difference of 25% in the proportions of adequate scores between two independent groups, with a power of 80% and an alpha of 0.05 (two-sided), we needed 20 IPs in each arm of the RCT. Given the fact that each IP made two cases at baseline and at follow-up and each case was scored with six PIs, each IP produced 12 NA/A-scores at both baseline and follow-up.

### Statistical Analysis

The baseline characteristics of the IPs in the two groups (CG and IG) were compared using crosstabs (Chi-square) for the categorical variables and independent T-test for the continuous variables.

Univariately, the outcome measures were analyzed with T-test for the scale scores of the PIs and the sum scores of the knowledge test. These tests were paired for the difference between T0 and T1 per group (CG or IG), and independently for the difference between groups (CG and IG) at T0 and T1. For the binary outcome measure the differences between the groups (CG and IG) at the two time points (T0 and T1) were analyzed with crosstabs (Chi-square).

The scale outcome variable of the PIs (1 to 7) was analyzed using linear mixed models. Besides the effect of Group, Time and their interaction effect on the outcome variable (the scaled PI score), we also corrected for the possible effects of the case, of the pair of Test-IPs, and of the kind of PIs per case on the PI scores of the participants. In our model fixed effects were: Intercept, Group (CG, IG), Time (T0, T1), Pair of Test-IPs (1, 2, 3), Group\*Time, Case (B, C, D, E) within Time, and PI (1 ... 6) within Case within Time. Besides these fixed effects, a random coefficient for the intercept with as 'subject' the IPs (1 ... 40) was calculated to account for possible clustering of the scores at IP level.

The binary outcome variable for the PI-score NA versus A was analyzed using Generalized Estimating Equations (GEE) with a logit link function, and with the same fixed effects as was the case in the linear mixed models analysis of the scale outcome variable, and for subject effect of the IPs (1 ... 40). Ancillary analyses were performed, using the differences in the estimated marginal means, to analyze the influence of case, pair of Test-IPs and PIs on the proportion of adequate scores.

For the knowledge test, the sum score of good answers per group before and after the training was calculated and analyzed using ANCOVA, with the sum score of the



knowledge test at T1 as dependent variable and the sum score of the knowledge test at T0 and Group as independent variables.

In the multivariate analyses possible confounding effects of background variables - of which the baseline characteristics of the IPs in the two groups (CG, IG) were significantly different - on the outcome variable were taken into account, i.e. a change of the coefficient of the variable Group of 10% or more. If so, a possible interaction effect of this confounding variable with the variable Group on the outcome variable was analyzed as well. All analyses were performed using SPSS 15.02.

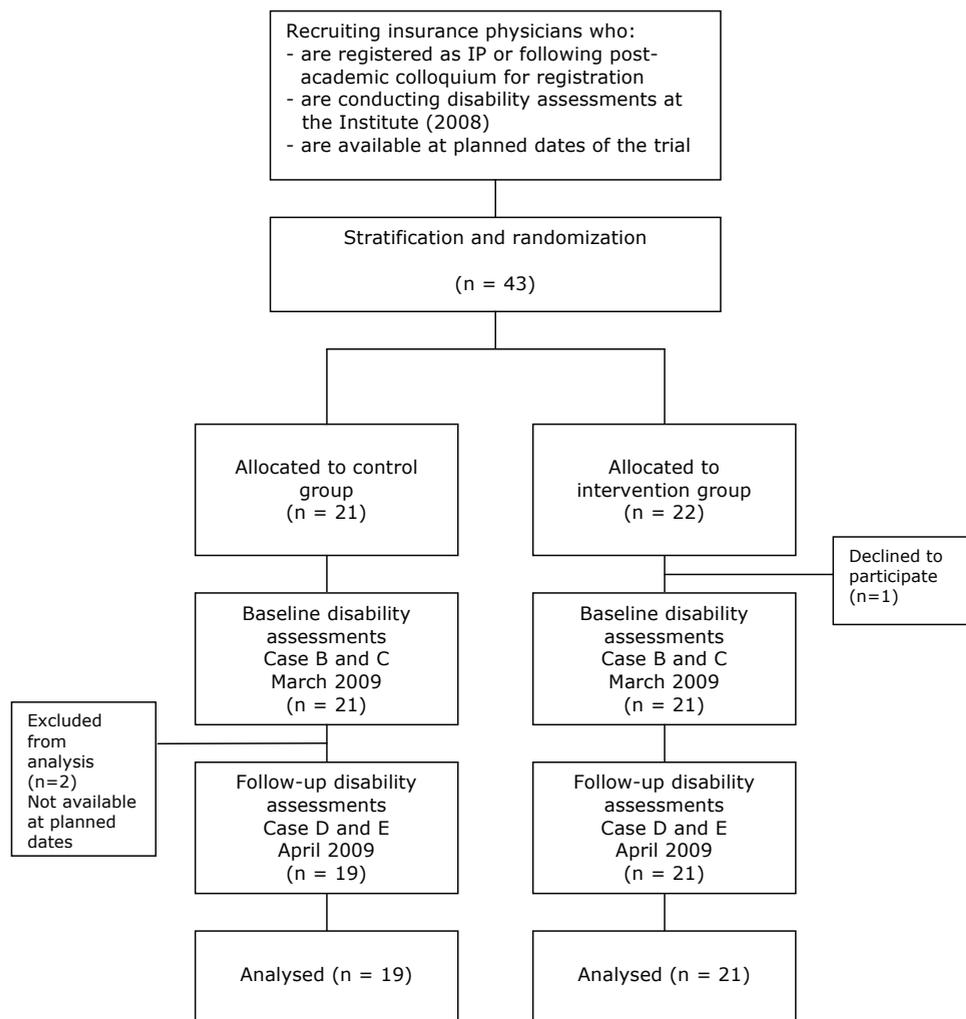
## Results

### Participant flow

Between January and March 2009 43 IPs applied for the course at the NSPOH. After the stratification and the randomisation procedure 21 were allocated to the usual care group and 22 to the intervention care group. One of the IPs who was allocated to the intervention group withdrew from the course and was lost to follow-up. Two IPs who were originally allocated to the control group were not available on the planned dates for the control group; they participated for reasons of education in the intervention group but were excluded from the RCT. Therefore the control group consisted of 19 participants and the intervention group of 21 (See the flowchart in Figure 5.1). The 40 IPs completed the trial and all data were obtained, except for the knowledge test results of two IPs of the 19 IPs from the usual care group. These two IPs refrained from this part of the measurements for personal reasons.

### Baseline characteristics

Baseline characteristics revealed a significant imbalance between the control group and the intervention group only for the variable 'mean number of clients with depression assessed by an IP per month'. If necessary, the results of all the analyses were adjusted for this variable. The other baseline characteristics (i.e. age, gender, being registered as an IP etc; see Table 5.2) were almost equally distributed across both groups (Chi-Square tests and T-tests were not significant). All participating IPs were currently conducting disability assessments at time of the RCT.



**Figure 5.1:** Flow chart of the participants through the phases of the randomised controlled trial. IP=Insurance physician; Institute=Dutch Institute for Employee Benefits Schemes.

## Outcomes

In the controlled setting, the guideline adherence of the group trained IPs ( $n = 21$ ) in the assessment of clients with depression (as measured with the scale outcome) increased by 16% compared with their guideline adherence at baseline, while the control group ( $n = 19$ ) showed an 8% decrease in their guideline adherence at follow-up. In the controlled setting, the guideline adherence of the trained IPs as measured with binominal outcome increased by 20%-points to 71%, while the guideline adherence of the control group decreased by 5%-points to 43%. The outcomes per different kinds of analysis are given in Tables 5.3, 5.4 and 5.5, and are illustrated below.

**Table 5.2:** Baseline characteristics of insurance physicians in control group (CG) and intervention group (IG).

Baseline characteristics	CG (n = 19)	IG (n = 21)
	Mean (sd) or percentage	
Age in years	50.5 (6.7)	51.1 (6.2)
Male	47%	52%
Weekly working hours	31.8 (9.9)	31.1 (9.2)
Years working as physician	21.7 (6.4)	23.5 (5.1)
Registered as insurance physician	84%	86%
Years working as insurance physician	15.4 (8.1)	15.6 (7.9)
Intensity of kind of professional activities	4.1 (0.8)	3.9 (0.8)
Number of clients with depression assessed per month*	9.3 (5.6)	5.3 (3.7)
Assessment time for depressed clients (minutes)	136.3 (62.3)	153.7 (48.4)
Assessments under the new disability act	68%	52%
Employee of the Institute	79%	81%
Attitude to guidelines in general (Scale 9-45)	30.8 (5.4)	31.7 (6.8)
Attitude to the GD (Scale 9-45)	31.8 (4.1)	33.9 (6.2)
Intention to use the GD (Scale 10-50)	34.5 (5.5)	35.0 (6.0)
Use (self-reported) of the GD (Scale 1-5)	3.1 (1.2)	3.0 (1.1)

\*Significant difference between both groups, possible confounder; GD=Guidelines for Depression; Institute=the Dutch Institute for Employee Benefits Schemes.

The bivariate analyses of the data showed significantly higher scores for the intervention group at T1. The paired T-Test of the PI score between T0 and T1 was not significant for the control group ( $p = 0.092$ ), but was significant for the intervention group ( $p < 0.0005$ ). The unpaired T-Test of the difference between control group and the intervention group was not significant at T0 ( $p = 0.32$ ) and significant at T1 ( $p < 0.0005$ ). The crosstabs revealed similar results for the (percentages of) Adequate scores (See Table 5.3).

The unpaired T-test of the second outcome, the knowledge of the IPs of the guidelines for depression, showed no significant differences between groups at baseline ( $p = 0.28$ ) and was significant at follow-up ( $p = 0.006$ ). The paired T-test between baseline and follow-up was not significant for the control group ( $p = 0.84$ ) and significant for the intervention group ( $p = 0.017$ ).

The multivariate analyses of the data resulted in statistically significant differences between groups for the IPs' performances on applying the guidelines for depression. The mixed models analysis produced a significantly higher score at the PI-scale score,

i.e. guideline adherence, for the intervention group compared with the control group, accounted for possible effects of variables at different levels such as the pair of Test-IPs, the case or a certain PI used within a case (see Table 5.4). The GEE analysis, which also accounted for the effect of the same variables at the various levels, showed similar results for the binomial PI score (results not shown here). The estimated marginal means of the guideline adherence (NA/A-score) per PI and per case showed that the trained IPs (IG) performed significantly better at each PI and at both cases at follow-up (results not shown here).

**Table 5.3:** Outcome measures control group (CG) and intervention group (IG) at T0 and at T1 and p-values between-group differences at T1.

	CG (19 physicians)		IG (21 physicians)		IG-CG Difference	
	T0	T1	T0	T1		
<b>Guideline adherence IPs to GD</b>	<b>228 scores</b>		<b>252 scores</b>			<b>p-value</b>
- Mean (SD) PI-sumscores (1-7)	3.6 (1.9)	3.3 (1.9)	3.8 (1.9)	4.4 (1.6)	T-Test	p < 0.0005
- Adequate scores (%)	48%	43%	51%	71%	Crosstabs	p < 0.0005
	CG (17 physicians)		IG (21 physicians)		IG-CG Difference	
	T0	T1	T0	T1		
<b>Knowledge test IPs on GD</b>						
- Mean (SD) test scores (0-10)	5.1 (1.2)	5.1 (1.3)	5.5 (1.4)	6.3 (1.2)	T-Test	p = 0.006

**Table 5.4:** Results of the Mixed Models analysis of the primary outcome guideline adherence.

	Group	Estimated means		Interaction effect (se) p-value
		T0	T1	
<b>PI-scale score</b> (1 to 7)	<b>CG</b>	3.62	3.32	0.97 (0.19)
	<b>IG</b>	3.77	4.44	p < 0.0005

Difference of Performance Indicator (PI) scale score outcomes between the insurance physicians (IP) in the Intervention group (IG) and the Control group (CG) at time T0 and T1. The estimated means and the interaction effect (time\*group) with standard error (se) are presented. Mixed Models Analyses: adjusted for fixed effects of Pair of test-IPs (1, 2, 3), Case (B, C, D, E) within Time, PI (1 ... 6) within Case within Time and possible influences of clustering on the level of insurance physicians.



The ANCOVA analysis of the secondary outcome knowledge test sum score produced significantly higher scores for the intervention group compared with the control group. The results of the ANCOVA analysis had to be adjusted for the confounding variable 'mean number of clients with depression, assessed by an IP per month' (see Table 5.5).

**Table 5.5:** Results of the Ancova analysis of secondary outcome knowledge.

	Group	Estimated means	
		T1 <sup>(1)</sup>	T1 <sup>(2)</sup>
<b>Knowledge Test Sum Score</b> (0-10)	<b>CG</b>	5.24	5.29
	<b>IG</b>	6.19	6.15
<b>Group effect (se) p-value</b>		0.95 (0.36) 0.012	0.86 (0.40) 0.038

Difference of Knowledge Test Sum Score outcomes between intervention group (IG) and control group (CG) at T1. The estimated means, and the Group effect with standard error are presented.

<sup>(1)</sup> Adjusted: covariate 'Knowledge Test Sum Score at T0' was evaluated at the mean value.

<sup>(2)</sup> Adjusted: covariate 'Knowledge Test Sum Score at T0' and 'Mean Number of Clients with Depression Assessed by an IP per Month' were evaluated at their mean values.

## Discussion

### Main findings and interpretation

Due to the newly developed implementation strategy, gains in guideline adherence and knowledge were obtained in the controlled setting of this study. The results of the knowledge test between the groups at follow-up were significant, although the difference was smaller than that for guideline adherence.

In the controlled setting of this study, trained IPs performed their work disability assessments more in concordance with the guidelines than did those from the control group. It appeared that the trained IPs produced significantly more adequate scores in all six PIs. With these scores they distinguished themselves from the control group in each of the main points of the guidelines, as measured by the PIs at follow-up.

After adjustment for confounding, the result of the knowledge test showed a smaller positive effect for intervention. A logical explanation for this is that fewer gains in knowledge of the guidelines can be achieved, due to the intervention, by IPs who are already more familiar with the disorder, suggesting that the greater the number of clients with depression an IP assesses per month, the more their knowledge of this disorder



increases. Given the outcomes of the control group, it should be remarked that we found no indications for a learning effect from the measurements (assessing the work disability of the clients), because the guideline adherence in the control group seemed to decrease slightly in the follow-up measurement.

The scores of correct answers on the knowledge test were rather low for both groups at baseline and at follow-up, but increased significantly for the intervention group compared with the control group at follow-up. However, having knowledge of the guidelines does not imply application of the guidelines in practice. Guideline adherence as a concept should be regarded as behaviour involving more than knowledge. Furthermore, the main goal of training was that the IPs practised their skills in applying the guidelines, and not specifically that they improved their knowledge; the improvement in knowledge can be considered as a desirable side effect of the training.

### Strengths and limitations

The strength of this study was that the intervention implementation strategy was developed on the basis of the needs of IPs and the opinions of experts [8]. Another strong point was that we were able to measure quality with a valid and reliable instrument, the developed PIs [14]. In the analyses we accounted for the possible effects of factors at different levels on the outcomes, such as the influence of the case, the pair of Test-IPs and the PI itself. The results were confirmed in each of the different types of analysis. The design of this RCT, in which four different clients were each simultaneously assessed by groups of IPs, provided results that allowed us to draw sound and specific conclusions concerning the efficacy of the intervention. The developed implementation strategy improved guideline adherence in a controlled setting.

However, this design had limitations as well. We could not evaluate the effectiveness of the implementation strategy in practice, and thus the results of this study cannot be directly translated into practice. For practical reasons, we had to conduct the RCT in a fixed laboratorial setting. In this RCT the disability assessments were performed under specific conditions: the clients were presented on a video screen, the participating IPs could not speak to them but had to assess their disability based on the information presented by the actor playing the client on the screen. Nevertheless, this RCT was based on and translated from the situation of IPs working in practice, and thereby offered us the optimal conditions for studying the efficacy of our newly developed implementation strategy. Another limitation of this study is the short time line of the RCT. This RCT contained one follow-up measurement, so that long-term effects of the training could not be evaluated; however, the justification for this short time line in the design of the RCT was the risk of contamination between groups, i.e. the possibility that the control group may be contaminated by the intervention group in the period after the intervention group had received the intervention programme. We therefore planned the intervention programme immediately after the follow-up measurement of the control

group, making contamination of the control group by the intervention group impossible. The overall time interval between the start and the end of the RCT was no more than two weeks. By shortening the timeline of the trial, we limited the risk of influences from outside the laboratorial setting. Furthermore, selection bias of the IPs who participated in this study is possible since they participated voluntarily.

### Comparison with literature

The effectiveness of continuing medical education with the aim of stimulating physicians' guideline adherence has been evaluated for clinical care, primary care and occupational health care [5, 16-19]. The resulting overview was that most effects could be expected from multifaceted interventions characterized by mixed educational programmes with an active role for the physicians. Although in this RCT, where guideline adherence increased from 51% to 71%, this expectation was confirmed for the field of insurance medicine, our implementation strategy still has to be evaluated in practice where there may be more barriers to implementation than the physician's behaviour. Two primary care studies concerning multifaceted interventions in the implementation of guidelines did not demonstrate any increase in guideline adherence in practice [18, 20]. In another study in primary care, multifaceted intervention strategy only modestly improved implementation of guidelines for low back pain [21]. The overall adherence rate to 70 guidelines in primary care within a period of 10 years was 67% [2]. For mental health disorders the figures of guideline adherence were even lower. In occupational health care, guideline adherence of 39% was found for mental health problems [22], while in primary care guideline adherences for depression and for anxiety disorders of 42% and 27% respectively were reported in a cross-sectional study [23]. Our findings showed that with a multifaceted strategy for a mental health disorder such as depression sizeable gains in guideline adherence could be achieved in a controlled setting. We can add to the conclusions of another insurance medicine study [24], in which it was found that a workshop improved evidence-based skills and self-efficacy, that an evidence-based medicine approach can be successfully adapted to the field of insurance medicine.

### Practice implications

The educational intervention was evaluated in this study with the participation of a limited group of IPs. Now that the efficacy of this training has been shown in the controlled setting of this study, distribution to other IPs is recommended. Though this training was developed for the guidelines for depression, with adjustments it could be used for the implementation of other insurance medicine guidelines as well. The results of the developed implementation strategy did not show evident barriers at the level of the physician. Potential organizational barriers such as available time for a physician to apply guidelines in practice could not be investigated in this RCT. However, this training programme can be seen to suit the needs of physicians and their employers, because



its implementation requires the investment of only one day of the physician's time, while the return proves to be relatively high. The RCT was carried out in a controlled setting for practical reasons (i.e. to have groups of IPs assessing the same video-taped clients simultaneously), but the implementation strategy (the training programme with evidence based-tools) itself is ready to be carried out in the real world setting. Finally, clients are expected to benefit from 'guideline proof' assessments, because the quality of these assessments is higher.

### Conclusions

In this study the efficacy of a newly developed implementation strategy for the insurance medicine guidelines for depression was evaluated in an RCT. In a controlled setting, the implementation strategy did improve the guideline adherence of IPs in the disability assessments of clients with depression and gains in knowledge concerning the guidelines were achieved. Though the guideline adherence of the trained IPs increased sizeably under the specific conditions in this study, it is yet not known whether these effects will be retained in the long term. Therefore, further research on the long-term effectiveness of this educational intervention is needed. This educational intervention is suitable for practice, because it combines a high success rate with low investment as the training takes only one day of the physician's time.

### Competing interests

The authors declare that they have no competing interests. Funding sources: FZ, JRA and AJMS are (partially) funded by UWV. The study sponsor had no role in the study design, in the collection, analysis or interpretation of the data, in the writing of the case-reports or in the decision to submit the paper for publication. The design of this study was laboratorial and for data collection fictitious but realistic case-reports were used. Consequently, the Medical Ethics Committee agreed with the design.

### Authors' contributions

The authors declare that they participated in the study and made the following contributions to the study, and that they have seen and approved the final version. AJMS, FZ, JRA, DLK and AJvdB contributed to the conception and design of this study. AJMS, FZ and DLK contributed to the analysis. FZ and AJMS wrote the manuscript. JRA, DLK and AJvdB revised and commented on the manuscript. AJMS, JRA and AJvdB will act as guarantors of this study. The corresponding author (AJvB) had full access to all data in the study and had final responsibility for the decision to submit for publication.

### Acknowledgements

The authors wish to thank the insurance physicians who participated in this research. The Research Center for Insurance Medicine AMC-UMCG-UWV-VU University Medical Center,

in Amsterdam, is a joint initiative of the Academic Medical Center (AMC), the University Medical Center in Groningen (UMCG), the Dutch Institute for Employee Benefit Schemes (UWV), and the VU University Medical Center (VUMC).

### References

1. Grol R, Grimshaw J: From best evidence to best practice: effective implementation of change in patients' care. *Lancet* 2003, **362**:1225-30.
2. Grol R: Successes and failures in the implementation of evidence-based guidelines for clinical practice. *Med Care* 2001, **39**:46-54.
3. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al.: Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 1999, **282**:1458-65.
4. Grol R, Wensing M: What drives change? Barriers to and incentives for achieving evidence-based practice. *Med J Aust* 2004, **180**:57-60.
5. Davis DA, Thomson O'Brien MA, Freemantle N, Wolf FM, Mazmanian P, Taylor-Vaisey A: Impact of Formal Continuing Medical Education: Do Conferences, Workshops, Rounds, and Other Traditional Continuing Education Activities Change Physician Behaviour or Health Care Outcomes? *JAMA* 1999, **282**:867-74.
6. Bijl RV, Ravelli A, Van Zessen G: Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Soc Psychiatry Psychiatr Epidemiol* 1998, **33**:587-95.
7. Murray CJ, Lopez AD: Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet* 1997, **349**:1498-1504.
8. Zwerver F, Schellart AJM, Anema JR, Rammeloo KC, Van der Beek AJ: Intervention mapping for the development of a strategy to implement the insurance medicine guidelines for depression. *BMC Public Health* 2011, **11**:9.
9. Berkhof M, Van Rijssen HJ, Schellart AJ, Anema JR, Van der Beek AJ: Effective training strategies for teaching communication skills to physicians: an overview of systematic reviews. *Patient Educ Couns* 2011, **84**:152-62.
10. Bero LA, Grilli R, Grimshaw JM, Harvey E, Oxman AD, Thomson MA: Closing the gap between research and practice: an overview of systemic reviews of interventions to promote the implementation of research findings. *BMJ* 1998, **317**:465-8.
11. Davis DA, Taylor-Vaisey A: Translating guidelines into practice. A systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines. *Can Med Assoc J* 1997, **157**:408-16.
12. Health Council: *Assessment, treatment and counselling. Medical responses to sickness absenteeism en occupational disability*. Den Haag: Gezondheidsraad, 2005.
13. Health Council: *Insurance medicine guidelines for depression*. Den Haag: Gezondheidsraad, 2006.
14. Schellart AJM, Zwerver F, Anema JR, Knol DL, Van der Beek AJ: Development and reliability of performance indicators for measuring adherence to a guideline for depression by insurance physicians. *Disabil Rehabil* 2011, **33**:2535-43.
15. Hamilton M: A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960, **23**:56-62.

16. Smits PBA, Verbeek JHAM, Van Dijk FJH, Metz JCM, Ten Cate ThJ: Evaluation of a postgraduate educational programme for occupational physicians on work rehabilitation guidelines for patients with low back pain. *Occup Environ Med* 2000, **57**:645-646.
17. Coomarasamy A, Taylor R, Khan KS: A systemic review of postgraduate teaching in evidence-based medicine and critical appraisal. *Med Teach* 2003, **25**:77-81.
18. Forsetlund L, Bradley P, Forsen L, Nordheim L, Jamtvedt G, Bjorndal A: Randomised controlled trial of a theoretically grounded tailored intervention to diffuse evidence-based public health practice. *BMC Med Educ* 2003, **3**:2.
19. Sanci LA, Coffey CMM, Veit FCM, Carr-Gregg M, Patton GC, Day N, Bowes: Evaluation of the effectiveness of an educational intervention for general practitioners in adolescent health care: randomised controlled trial. *BMJ* 2000, **320**:224-30.
20. Walldorff FB, Almind G, Mäkelä, Moller S, Waldemar G: Implementation of a clinical dementia guideline. A controlled study on the effect of a multifaceted strategy. *Scand J Prim Health Care* 2003, **21**:142-147.
21. Engers AJ, Wensing M, Van Tulder MW, Timmermans A, Oostendorp RAB, Koes BW, Grol R: Implementation of the Dutch Low Back Pain Guideline for General Practitioners: a Cluster Randomised Controlled Trial. *Spine* 2005, **30**:595-600.
22. Rebergen D, Hoenen J, Heinemans A, Bruinvels D, Bakker A, Van Mechelen W: Adherence to mental health guidelines by Dutch occupational physicians: *Occup Med* 2006, **56**:461-8.
23. Smolders M, Laurant M, Verhaak P, Prins M, Van Marwijk H, Penninx B, Wensing M, Grol R: Adherence to evidence-based guidelines for depression and anxiety disorders is associated with recording of the diagnosis. *Gen Hosp Psych* 2009, **31**:460-469.
24. Kok R, Hoving JL, Verbeek JH, Schaafsma FG, Smits PBA, Van Dijk FJH: Evaluation of a workshop on evidence-based medicine for social insurance physicians. *Occup Med* 2008, **58**:83-87.



## Appendix 5.1: Intervention implementation

### Intervention goals

To improve the IP's knowledge and skills and self-efficacy in applying the guidelines for depression. Practice reinforcement of the IP's assessments of clients with depression.

Learning objectives for the participating IPs:

- learn to perform disability assessment in concordance with guidelines for depression
- learn to apply tools to recognize depression and to assess the working ability of claimants with depression
- learn to base conclusions in the disability assessments on convincing arguments
- learn to write powerful, more transparent reports

### Intervention content

Explanation of the evidence based content of the guidelines for depression by an IP trainer. Translation of the guidelines for depression by IP trainer into six main elements, which are relevant for use in daily practice of the IP: diagnostics DSM IV, severity of the disorder, course, risk factors, co-morbidity, judgement of treatment and therapy, assessment of work ability. Disability assessment of a client with depression presented on video. Writing assessment report on the client with depression.

### Educational strategy

#### Course manual

Learn objectives, list of used literature, suggestions for further reading. The complete guidelines of depression in book format. Resume of the common principles of reasoning and in particular applied for the guidelines for depression (with practical examples). Tools with the aim of making the guidelines easier to use in practice: Desk mat, summary with main elements from the guideline/evidence based medicine, checklists, Hamilton Depression Rating Scale [15] Handouts of presentations by psychiatrist and the trainer IP.

#### One day training

Debriefing from the baseline measurements. Taking a knowledge test concerning the guidelines for depression. Inter-active presentation by the trainer IP concerning the guidelines for depression. The trainer translates the guidelines into the insurance physician's daily practice. Client with depression played by an actor is presented on video to the group. Workshop with IPs in subgroups learning to use practical tools for the assessment of the client with depression. Presentations of the findings per subgroup to the complete group, with feedback from the IP trainer. Modelling of written arguments in the assessment reports about the client with depression by the trainer IP. Evaluation of the training and taking the same knowledge test as at the start of the training.

IP=Insurance physician.

# Chapter

---

# 6

## **Changes in insurance physicians' attitudes, self- efficacy, intention, and knowledge and skills regarding the guidelines for depression, following an implementation strategy**

*J Occup Rehabil* 2012, accepted for publication

**F Zwerver  
AJM Schellart  
JR Anema  
AJ van der Beek**

## Abstract

**Introduction:** To improve guideline adherence by insurance physicians (IPs), an implementation strategy was developed and investigated in a randomized controlled trial (RCT). This implementation strategy involved a multifaceted training programme for a group of IPs in applying the guidelines for depression. In this study we report the impact of the implementation strategy on the physicians' attitude, intention, self-efficacy, and knowledge and skills as behavioural determinants of guideline adherence. Any links between these self-reported behavioural determinants and levels of guideline adherence were also determined.

**Methods:** Just before and three months after the implementation of the multifaceted training, a questionnaire designed to measure behavioural determinants on the basis of the ASE (Attitude, Social Norm, Self-Efficacy) model was completed by the intervention (n=21) and the control group (n=19). Items of the questionnaire were grouped to form scales of ASE determinants. Internal consistency of the scales was calculated using Cronbach's alphas. Differences between groups concerning changes in ASE determinants, and the association of these changes with improvements in guideline adherence, were analyzed using analysis of covariance (ANCOVA).

**Results:** The internal consistency of the scales of ASE determinants proved to be sufficiently reliable, with Cronbach's alphas of at least 0.70. At follow-up after three months, the IPs given the implementation strategy showed significant improvement over the IPs in the control group for all ASE determinants investigated. Changes in knowledge and skills were only weakly associated with improvements in guideline adherence.

**Conclusions:** The implementation strategy developed for insurance physicians can increase their attitude, intention, self-efficacy, and knowledge and skills when applying the guidelines for depression. These changes in behavioural determinants might indicate positive changes in IPs' behaviour towards the use of the guidelines for depression. However, only changes in knowledge and skills related to the use of the guidelines were associated with improvements in IPs' actual performance when applying the guidelines.



## Introduction

Health care guidelines are intended to incorporate evidence-based medicine into the daily practice of physicians [1-3]. Encouraging the use of guidelines in daily practice is important for improving uniformity and quality in health care. However, the implementation of guidelines is a complex process influenced by many factors, such as the behaviour of the physicians, the guidelines themselves or the way in which the guidelines are implemented [4, 5]. There are numerous possible barriers in this process, which range from the distribution of guidelines to the use of guidelines in practice by physicians. Such barriers can be external, such as lack of availability, lack of practical relevance of the guidelines, or lack of support by the organization. But barriers can also be internal, for example lack of familiarity with the guidelines, lack of physicians' agreement with guidelines, negative attitudes in general towards guidelines (some physicians refer to guidelines as 'cookbook medicine'), or lack of self-efficacy in using guidelines [5].

Physicians' behaviour towards using guidelines may well be one such barrier, and therefore requires investigation. Physicians give various reasons for their reluctance to use guidelines and these reasons include the following: guidelines do not suit the individual problems of their patients, using guidelines does not improve their work (so-called lack of outcome expectancy), using guidelines limits their professional independency, or there is no pressure from patients or staff to use guidelines. The challenge for educational interventions is therefore to positively influence physicians' behaviour towards the use of guidelines.

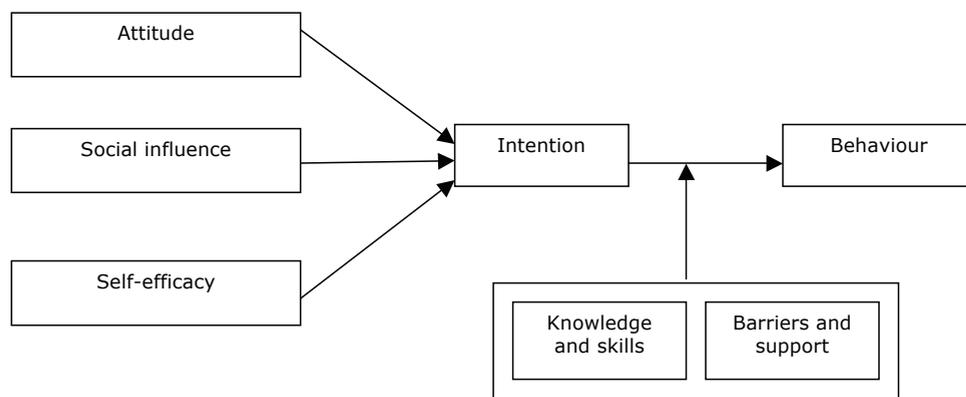
Educational programmes for physicians – as part of a guideline implementation strategy – have been evaluated in other studies and have produced varying results. Changing physicians' behaviour, such as increasing their guideline adherence, is possible, but such change requires comprehensive approaches at different levels, tailored to specific settings and persons [6]. With regard to the educational aspects of physicians' guideline adherence, the strongest effects can be expected from multifaceted interventions rather than more formal types of education such as stand-alone lectures [7].

In 2005, the Dutch Health Council implemented guidelines in the field of insurance medicine [8]. For one of these guidelines – the guidelines for depression – we subsequently set up a research project to evaluate a newly developed implementation strategy [9]. This implementation strategy consisted of a tailor-made training programme for insurance physicians (IPs) in which – facilitated with various tools – they learned to apply the guidelines for depression. The information in all tools was evidence-based, derived from the guidelines, and readily applicable in practice.

Explanations for physicians' behaviour with regard to guideline adherence can be found in the Attitude, Social Norm, Self-Efficacy model (ASE model), which is derived



from the Theory of Planned Behaviour (TPB) [10, 11]. The TPB is a theory designed to predict and explain human behaviour in specific contexts. Behaviour (applying guidelines) is influenced by intentions to perform that behaviour. In turn, these intentions are preceded by attitude, social norm and self-efficacy with regard to the desired behaviour. IPs are thought to have a certain attitude (positive or negative) towards guidelines that influences their intention to use them. Furthermore, IPs' intention to use guidelines could be determined by their colleagues (social norm) or by their perception of behavioural control, i.e. the degree to which they feel comfortable using guidelines (self-efficacy). The relationships between the determinants of behaviour – such as attitude, social norm, self-efficacy and intention – and the interfering stimuli or barriers involved in performing expected behaviour are shown in the ASE model (Figure 6.1).



**Figure 6.1:** The ASE model, as defined by De Vries [12].

We started by investigating and describing the behaviour of IPs on basis of the ASE model by means of a cross-sectional study at baseline [12]. In this study we investigated the relationships between the IPs' intention to use the guidelines for depression and their self-reported determinants of behaviour towards their use of the guidelines. However, the ASE model could only partly be confirmed. We found no relationship between intention and use of the guidelines while it is this relationship which represents the main line of the ASE model. We did however find determinants of behaviour that influenced the IPs' intention to use the guidelines, namely the influence of colleagues, self-efficacy and the way in which guidelines are implemented.

We subsequently carried out a randomized controlled trial (RCT) to evaluate the implementation strategy that we developed for the guidelines for depression, with guideline adherence by IPs as primary outcome [13]. This RCT showed an improvement in guideline adherence for the IPs given the implementation strategy.

Considering the limited findings with regard to the ASE model at baseline, but given the clear results of the RCT, we wondered what differences and changes

in ASE determinants might occur between the groups of the RCT as a result of the implementation strategy. The implementation strategy was predominantly aimed at training the IPs' skills and improving their self-efficacy by giving them practice in applying the guidelines with the help of tools, and in analyzing case studies of clients with depression played by actors presented on video. If the aim is to change behaviour, improving knowledge of the guidelines alone might not be sufficient. If so, it is important to know whether these differences or changes in ASE determinants could be related to the improvements observed in physicians' guideline adherence, and, above all, which of the ASE determinants might predict improvements in physicians' guideline adherence.

The aims of the present study are firstly to describe changes in determinants of IPs' behaviour following an implementation strategy for the guidelines of depression and secondly to investigate their association with the changes observed in guideline adherence.

The Medical Ethics Committee of VU University Medical Center approved the study design and the Netherlands Trial Registration accepted the RCT: NTR1863.

## Methods

### Participants

The participants were insurance physicians (IPs) that we recruited with the help of the Netherlands School of Public and Occupational Health (NSPOH) and who attended a post-graduate course in applying the insurance medicine guidelines for depression. The IPs participated on a voluntary basis. The RCT was integrated into this course. Inclusion criteria for enrolment of participants in the RCT was that they were 1) either registered as an IP, or still following the post-graduate course in Insurance Medicine, and 2) conducting disability assessments of clients under commission of the Dutch National Institute for Employee Benefits. In the Netherlands, the Institute for Employee Benefits (the Institute) is responsible for evaluating disability claims. Dutch employees (known by the Institute as clients) can claim disability benefits after having been sick-listed for 104 weeks, during which time the employees are attended to by occupational physicians. After 104 weeks the employees are transferred from the occupational physician to the IP, who assesses the work disability claim at the Institute. The level of the employee's benefit is then determined on the basis of this work disability assessment by an IP. An IP is a physician who has completed a four-year post-graduate course combined with practice as an IP, resulting in registration as an IP.

Forty-three IPs applied for the training on a voluntary basis. All 43 were included and they were individually allocated in order of registration either to the intervention group or to the control group using a random-sequence table. To prevent unequal allocation across groups, the participants were stratified before randomization according to three



prognostic factors: age, gender and registration as an IP. A research assistant performed the randomization and stratification procedure. The participants were all blinded for the design of the RCT and were given badges with a unique number, which was also written on their study materials. This meant that the researchers were also blinded for their group allocation. By the time it was clear whether participants were in the intervention or control group, due to the kind of training they received, the baseline measurements of the RCT had already been completed. The group of IPs was representative of the approximately 900 IPs working at the Institute [14]. The training programme was located at the NSPOH in Amsterdam. The participants went there on four different days from March to June 2009 for the measurements, the intervention programme and the control programme. Meanwhile they worked as practising IPs at the Institute.

### Study design

We conducted an RCT that compared an intervention group (IG) and a control group (CG) for guideline adherence and ASE determinants. Participants in the intervention group received the implementation strategy developed for the guidelines for depression. The control group received an alternative programme that did not interfere with the intervention programme. After completion of the RCT measurements, both groups followed the remaining training programme. Objective measurements regarding guideline adherence were carried out using performance indicators (PIs). The development and reliability of these PIs have been reported previously [15], as have the results of the RCT based on the PI scores observed [13]. The minimum sample size required to detect a change in the primary outcome of the RCT – guideline adherence – was determined by means of power analysis [13].

Another part of the same RCT measured ASE determinants using two questionnaires, one before the intervention and the other after the intervention. The ASE determinant ‘social norm’ was left out of this study, because changes in social norm are beyond the scope of the intervention.

### Questionnaires

Questionnaire constructs for measuring the four determinants attitude, self-efficacy, intention, and knowledge and skills were developed from the concepts in the ASE model [11]. The questionnaires at baseline (T0) and at follow-up after three months (T1) both contained identical constructs for the ASE determinants (Attitude, Self-efficacy, Knowledge, and Intention). The questionnaires used 38 items with responses on a five-point scale ranging from ‘strongly disagree’ to ‘strongly agree’. The items were clustered in four scales for the constructs determined by the ASE determinants (see Appendix for the questionnaires).

### Intervention

The implementation strategy was developed with the help of users and experts. It was aimed at improving the availability and the practicability of the guidelines for depression. The intervention consisted of a multifaceted training programme for IPs in applying the guidelines for depression. The different components of the programme included interactive presentations by experts and exercises in subgroups, where IPs practised assessing clients with depression played by actors on video. The IPs’ trainers provided them with feedback. Individual assignments for IPs involved practice in writing disability reports following the feedback from the IP trainers. A number of evidence-based tools (a plastic desk mat listing a summary of the guidelines, as well as two different guidelines checklists and the Hamilton Rating Scales for Depression [16]) were developed for this programme, aimed at improving the applicability of the guidelines. The IPs were instructed how to use these tools. Learning objectives for the IPs were to use the tools to improve their diagnostic skills, to improve their assessment of work ability of clients with depression, and to write down their findings and conclusions in well-argued reports. The objective for the IPs’ reports was that they should be more transparent and more evidence-based and should contain well-argued assessments of a client’s work ability. Two of the authors (FZ and JRA) acted as the two IP trainers in the intervention programme.

### Control group

For reasons of recruitment and equal treatment, the control group acted as a waiting list control and was later given the same educational programme as the intervention group. At the same time the intervention group received the intervention programme, the control group participated in a ‘placebo training’, which was a programme in motivational interviewing. The content of the motivational interviewing programme did not interfere with the intervention or the guidelines for depression, because these two programmes shared no common ground.

### Data collection and outcome measures

Data were collected using two questionnaires at baseline (T0) and at follow-up after three months (T1). The first questionnaire included items for the baseline characteristics of the participating IPs. The questionnaires were filled in and collected while participants attended the course. The primary outcomes were the IPs’ behavioural changes (T1 versus T0) towards the guidelines, expressed in terms of ASE determinants. We also determined the association, if any, between these self-reported ASE determinants and the main outcome of the RCT, i.e. the IPs’ levels of guideline adherence expressed in terms of a performance indicator (PI) sum score.

## Statistical analysis

The RCT required equal allocation to both groups of the participant characteristics age, gender and registration as an IP. If necessary, we corrected for confounding variables in the analyses performed. Both questionnaires contained items that formed constructs representing the four scales of the ASE determinants. Scale scores were obtained by adding the responses to the items within each scale. The internal consistency within each scale was determined with Cronbach's alphas. A Cronbach's alpha of 0.60 at baseline was considered to be the minimum for consistency. To investigate the differences between the groups due to the intervention, a one-way Analysis of Covariance (ANCOVA) was conducted ( $p < 0.05$ ). The group was the independent variable, the ASE determinant at baseline (T0) was the covariate, and the ASE determinant at follow-up (T1) was the dependent variable. All dependent variables were normally distributed, and the homogeneity-of-slopes assumption was tested for these variables – both conditions for the valid use of ANCOVA analysis. To address the second aim of this study, we used an ANCOVA model for change [17] with as dependent variable the PI sum score at T1. As independent variables we used as factor the intervention group and control group; and as covariates we used the PI sum score at T0, the relevant ASE determinant at T0, and the interaction between the change of the same ASE determinant at T1 and the group. We were especially interested in this interaction effect.

In all ANCOVA analyses we corrected for a confounding variable seen in the participating IPs: there was a significant difference between the intervention group and control group in the mean number of clients with depression that they assessed for work disability each month.

The statistical analyses were performed at the individual level of the participants in the RCT according to the per-protocol principle and using SPSS version 15.0. Because the trial in this study was an efficacy trial, in which we were interested in knowing whether the intervention works for a group of IPs in a specific controlled setting, rather than an effectiveness study carried out in real practice, we chose to present the figures of the per-protocol analyses. We also performed an intention to treat analysis, thereby including the three IPs lost to follow up (21 in the intervention group, and 22 in the control group), but this had no influence on the results.

## Results

A total of 43 IPs applied to take part in the post-graduate course: 21 received the intervention programme while 19 were in the control group. Three IPs were lost to follow-up. The response rate of the questionnaires was 98%. The baseline characteristics are summarized in Table 6.1. The IPs' behaviour regarding the guidelines for depression as determined by the ASE variables, which was the primary outcome measure, was related to the mean number of clients with depression per month assessed by the participating IPs.



**Table 6.1:** Baseline characteristics of the participants of the post-graduate course.

	Intervention	Control (n=21)	P-value* (n=19)
Years of age	51.1 (6.2)	50.5 (6.7)	0.92
Working hours/week	31.1 (9.2)	31.8 (9.9)	0.82
Years of experience as IP	15.6 (7.9)	15.4 (8.1)	0.92
Mean number of clients with depression per month	5.3 (3.7)	9.3 (5.6)	0.01
Gender M/F (%)	52.3 / 47.7	47.3 / 52.7	0.75
Being registered as IP (%)	85.7	84.2	0.89

Means and standard deviations are given for continuous variables. Percentages are given for the categorical variables. \*P-value of the independent t-test and P-value of the Chi-squared, respectively, between the intervention and the control group; significant differences are bold.

## Reliability

For each of the self-reported scales used for the ASE determinants, Cronbach's alphas were calculated to test their internal consistency. All ASE determinants had a Cronbach's alpha of at least 0.70, which indicates that the self-reported measures utilized in this study were sufficiently reliable (see Table 6.2).

**Table 6.2:** Internal consistency analysis (n=40).

Variables based on determinants of the ASE-model	Items	Cronbach's alpha Baseline (T0)	Cronbach's alpha Follow-up (T1)
Attitude (GD)	9	0.77	0.77
Self-efficacy	11	0.75	0.86
Knowledge and Skills	8	0.77	0.72
Intention	10	0.75	0.79

GD=Guidelines for Depression.

## Outcomes

Table 6.3 shows, for each of the ASE variables, the means, the standard deviations at baseline and at follow-up, and the p-values of the between-group difference at follow-up, corrected for baseline values and for the confounding variable 'mean number of clients with depression assessed per IP per month' ( $p < 0.05$ ). The intervention had a significant effect on all the ASE variables investigated. At follow-up the participants of the intervention group had not only a more positive self-reported attitude, self-efficacy and intention towards the guidelines for depression than the participants in the control

group, they also showed an improvement in knowledge and skills for applying the guidelines. In the intervention group, attitude and intention both improved by 12%, self-efficacy by 10%, and knowledge and skills by 5%. In the control group, attitude and intention stayed almost equal, while self-efficacy and knowledge and skills even decreased by 9% and 15%, respectively.

**Table 6.3:** Mean scores (SD) on the ASE-variable scales and p-values of the differences between intervention group (IG) and control group (CG) at follow-up (T1).

ASE-variables (scale)	Intervention (n= 21)		Control (n= 19)		P-value*
	T0	T1	T0	T1	
Attitude (scale 9-45)	33.9 (6.2)	38.1 (4.7)	31.8 (4.1)	33.7 (3.2)	<b>0.003</b>
Self-efficacy (scale 11-55)	35.8 (5.1)	39.5 (5.5)	34.6 (5.9)	31.6 (4.9)	<b>0.000</b>
Knowledge and skills (scale 8-40)	27.9 (5.4)	29.3 (4.5)	27.7 (5.5)	23.3 (4.0)	<b>0.000</b>
Intention (scale 10-50)	34.9 (6.0)	39.2 (5.8)	34.5 (5.5)	34.4 (4.6)	<b>0.014</b>

\*Ancova analysis: p-values of differences between Control group and Intervention group at follow-up, corrected for baseline value and the confounding variable (number of clients with depression). Bold figures are significant.

The ANCOVA analyses for the second aim of this study showed no significant interaction effects between changes in ASE determinants at T1 and group for the guideline adherence observed at T1 (see Table 6.4). With regard to the ASE determinants, only the interaction effect of the changes in knowledge and skills showed a weak association ( $p=0.093$ ) with the improvement in guideline adherence observed at T1. At the group level, only the associations for the control group tended to be weakly significant and negative for self-efficacy ( $p=0.111$ ) as well as for knowledge and skills ( $p=0.078$ ). The changes in attitude and in intention were not related to the improvement in guideline adherence observed at T1 ( $p=0.950$  and  $p=0.741$ , respectively).



**able 6.4:** Associations between observed guideline adherence at T1 (PI sum scores) with change of ASE-variables (self-reported) at T1 regarding the guidelines depression for intervention group (IG,  $n=21$ ) and control group (CG,  $n=19$ )\*.

ASE-variables	Parameter	(P-value)
Group*Dattitude T1		(0.950)
Group=CG	0.006	(0.991)
Group=IG	0.147	(0.758)
Group*Dself-efficacy T1		(0.213)
Group=CG	-0.578	(0.111)
Group=IG	0.081	(0.758)
Group*Dknowledge&Skills T1		(0.093)
Group=CG	-0.672	(0.078)
Group=IG	0.114	(0.801)
Group*Dintention T1		(0.741)
Group=CG	0.001	(0.996)
Group=IG	-0.258	(0.455)

\*Ancova analysis: dependent variable = PI sum score at T1; D = difference of ASE Variable T1-T0; parameter estimates and their p-values, with group as factor and corrected for baseline values of PI sum score and of the concerning ASE-variable, and for the confounding variable (number of clients with depression).

## Discussion

### Main results

In this study we investigated the effect of a newly developed implementation strategy on insurance physicians' attitude, self-efficacy, intention, and knowledge and skills towards the guidelines for depression. After three months, IPs who participated in the training course demonstrated a more positive attitude to the guidelines for depression, a higher intention to use them, more self-efficacy, and more knowledge and skills in applying the guidelines for depression than their colleagues in the control group.

### Interpretation

Our results show that the implementation strategy had the most impact on the physician's attitude, self-efficacy, and their intention to apply the guidelines and less impact on the physician's knowledge and skills. According to the ASE model, attitude and self-efficacy are the precursors of intention, which in turn predicts behaviour, in this case the physicians' guideline adherence. Physicians who were confident about applying the guidelines, and who had a positive attitude, showed a higher intention to use the guidelines. If attitude, self-efficacy and intention increase, subsequently facilitated by knowledge and skills, then behaviour should change positively. Whether this is a clinically

relevant change should be studied in real practice on patient outcomes. This change might also be clinically relevant, because if IPs are more inclined to apply guidelines in practice, their work disability assessments might be more evidence-based, and their assessments might be executed more uniformly. The implementation strategy indeed resulted in a change of behaviour, as we saw in the outcomes of the RCT. And as we have previously shown, the trained IPs showed better guideline adherence than the IPs in the control group [13].

The changes observed in the behavioural determinants appeared to last for a period of at least three months after the training took place. The fact that the control group showed a decrease in self-efficacy and in knowledge and skills indicates that there was no stimulating effect as a result of the measurements themselves; in fact these behavioural determinants actually faded with time. Although we expected there to be associations between the changes in ASE determinants and the improvements in guideline adherence, this was only marginally confirmed in this study for the determinants knowledge and skills. It is however possible that our training programme did induce the changes in these ASE determinants, as well as the improvements in guideline adherence.

### **Strengths and limitations of this study**

A strength of this study was that the questionnaires were developed on the basis of a theoretical model (the ASE model), and that they proved to be sufficiently reliable, while the constructs of this model were adjusted for the specific context of insurance medicine. Another strength was the high response rate of the questionnaires (98%), which was thanks to the design of the study. In this design there was a follow-up measurement after three months, giving the participants in the intervention group the opportunity to put into practice what they had learned and practised during the training programme. Finally, it was possible to link the ASE determinants to the main outcome of the RCT – guideline adherence – thereby providing insight into IPs' behaviour towards guidelines.

A limitation of this study was the low number of participants (40) used. Another limitation might be the fact that we studied changes in separate ASE determinants of behaviour as a result of the implementation strategy, while we were previously not able to confirm the ASE model in a cross-sectional analysis. However, in the present study it was possible to link such changes in ASE determinants to improvements in the levels of guideline adherence. Although the relationships between the ASE determinants of IPs' behaviour at baseline were not strong, ASE determinants did change after we changed the behaviour of IPs by training them in applying the guidelines for depression – in fact all four ASE determinants changed significantly in the expected direction in this group. An explanation for this phenomenon might be that the intervention directly influences all ASE determinants. An additional explanation could be that the ASE model is a better fit when describing changes in behaviour, instead of exploring behaviour only at a single



point in time. However, the fact that we could only demonstrate changes in the separate ASE determinants, and not in the relationships between the ASE determinants as a result of the implementation strategy, might well be a methodological limitation.

In the present study, three different kinds of bias might have occurred, which might also be regarded as methodological limitations. Firstly, the IPs participated on a voluntary basis, which might have induced a selection bias. However, since both the intervention and control groups were vulnerable to this bias, it might have reduced the contrast between both groups with regard to outcomes. Secondly, a literature search that assessed trends in self-reported adherence of clinicians to practice guidelines demonstrated that self-reported adherence levels exceeded the objective levels, resulting in a median overestimation of adherence of 27% [18]. Potential overestimation of self-reported guideline adherence may also have occurred in our study, but could not negatively influence our results since this bias accounts for both groups. Finally, in the follow-up questionnaire the participants were asked to fill in items relating to the intervention. Since the intervention took place three months previously this made their answers vulnerable to recall bias – they might have forgotten relevant facts, or they might have interpreted facts differently. On the other hand, the three-month interval was needed in order for the participants of the intervention group to reflect on what they had learned in the training programme. Furthermore, during the three-month interval they had the opportunity to put the acquired knowledge and skills concerning the guidelines into practice.

### **Comparison with other studies**

Other studies have suggested that theoretical models such as the Theory of Planned Behaviour (TPB) and its derivative, the Attitude, Social Norm, Self-Efficacy (ASE) model, could help to identify ways of improving physician adherence or even to predict that behaviour [19, 20]. Recently, one study reported that the ASE model appears to be suitable for the description of the assessment behaviour of IPs [21]. The results of our study confirm this. The insurance physicians who received the implementation strategy demonstrated not only a higher level of guideline adherence [13], but also significant improvements in the determinants of their behaviour. The insurance physicians in our study increased their attitude, self-efficacy and intention in applying the guidelines, all determinants that are precursors for the intended behaviour, i.e. use of guidelines. The determinant knowledge and skills increased far less than the determinants attitude, self-efficacy and intention in our study.

Physicians' knowledge of guidelines alone however, seems not to lead to better guideline adherence, as others have also shown [22, 23]. Furthermore, multifaceted interventions such as our implementation strategy are known to improve attitudes and behaviour, while stand-alone teaching only improves knowledge [24]. A cross-sectional survey carried out among Flemish occupational health physicians showed that the majority of physicians had a positive attitude toward implementing guidelines, but

the physicians mentioned barriers in legislative framework, education and information structure [25]. Given our positive results, the newly developed implementation strategy may well have been successful in removing such barriers in education and information structure. The combination of the educational strategies used in the training programme together with the translation of the guidelines into practical and useful tools for the IPs was probably what stimulated the IPs' attitude, self-efficacy, knowledge and skills, and intentions regarding use of the guidelines for depression.

### Practical implications

The implications of the newly developed implementation strategy, consisting of a multifaceted training programme, are encouraging. The training programme itself took only one day of the physicians' time. Similar training programmes could also be developed for other guidelines. This programme suited the needs of the physicians and was linked to their daily practice through the use of realistic case histories on video, which simulated clinical practice and contained evidence-based medicine. The IPs were then able to apply this evidence-based medicine in daily practice and gain experience in applying the guidelines for depression. Implementation of guidelines was also facilitated by the use of various tools. Educational programmes aimed at improving guideline adherence should be aimed not only at gaining knowledge but also at practising skills.

### Conclusions

The newly developed implementation strategy significantly increased the levels of insurance physicians' attitude, self-efficacy, intention, and knowledge and skills with regard to their use of the guidelines for depression. These changes in determinants of behaviour might indicate positive changes in IPs' behaviour regarding their use of the guidelines for depression. The improvements were achieved following a multifaceted one-day training programme, and lasted for at least three months. Although the levels of IPs' guideline adherence improved after receiving the implementation strategy, this gain could only be related to increased levels of knowledge and skills. Improving knowledge and skills seems to be weakly related to the improvements in observed guideline adherence.

### Competing interests

The authors declare that they have no competing interests. Funding sources: FZ, JRA and AJMS are partially funded by UWV. The study sponsor had no decisive role in this study.



### Authors' contributions

The authors declare that they participated in the study and made the following contributions to the study, and that they have seen and approved the final version. AJMS, FZ, JRA and AJvdB contributed to the conception and design of this study. AJMS and FZ contributed to the analysis. FZ and AJMS wrote the manuscript. JRA and AJvdB revised and commented on the manuscript. AJMS, JRA and AJvdB will act as guarantors of this study. FZ is the corresponding author.

### Acknowledgements

The authors wish to thank the insurance physicians who participated in this research. The authors would like to thank J.W.R. Twisk PhD, Professor of Applied Biostatistics at VU University Amsterdam, for his advice on modelling change. The Research Center for Insurance Medicine AMC-UMCG-UWV-VU University Medical Center in Amsterdam, is a joint initiative of the Academic Medical Center (AMC), the University Medical Center in Groningen (UMCG), the Dutch Institute for Employee Benefits Schemes (UWV), and the VU University Medical Center (VUMC).

### References

1. Grol R: Successes and failures in the implementation of evidence-based guidelines for clinical practice. *Med Care* 2001, **39**:1146-1154.
2. Rebergen D, Hoenen J, Heinemans A, Bruinvels D, Bakker A, Van Mechelen W: Adherence to mental health guidelines by Dutch occupational physicians. *Occup Med* 2006, **56**:461-8.
3. Forsetlund L, Bradley P, Forsen L, Nordheim L, Jamtvedt G, Bjorndal A: Randomised controlled trial of a theoretically grounded tailored intervention to diffuse evidence-based public health practice. *BMC Med Educ* 2003, **3**:2.
4. Grol R, Wensing M: What drives change? Barriers to and incentives for achieving evidence-based practice. *Med J Aust* 2004, **180**:S57-S60.
5. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al.: Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 1999, **282**:1458-65.
6. Grol R, Grimshaw J: From best evidence to best practice: effective implementation of change in patients' care. *Lancet* 2003, **362**:1225-30.
7. Berkhof M, Van Rijssen HJ, Schellart AJ, Anema JR, Van der Beek AJ: Effective training strategies for teaching communication skills to physicians: an overview of systematic reviews. *Patient Educ Couns* 2011, **84**:152-62.
8. Health Council: *Assessment, treatment and counselling. Medical responses to sickness absenteeism and occupational disability*. Den Haag: Gezondheidsraad, 2005.
9. Zwerver F, Schellart AJM, Anema JR, Rammeloo K, Van der Beek AJ: Intervention mapping for the development of a strategy to implement the insurance medicine guidelines for depression. *BMC Public Health* 2011, **11**:9.
10. Ajzen I: The theory of planned behavior. *Organ Behavi and Hum Decis Process* 1991, **50**:179-211.

11. De Vries H, Dijkstra M, Kuhlman P: Self-efficacy: third factor besides attitude and subjective norm as a predictor of behavioral intention. *Health Educ Res* 1988, **3**:273-282.
12. Zwerver F, Schellart AJM, Anema JR, Van der Beek AJ: Relationships between intention to use guidelines and determinants of insurance physician's behaviour. Submitted, 2012.
13. Zwerver F, Schellart AJM, Knol DL, Anema JR, Van der Beek AJ: An implementation strategy to improve guideline adherence of insurance physicians: an experiment in a controlled setting. *Implement Sci* 2011, **6**:131.
14. Van Rijssen HJ, Schellart AJM, Anema JR, Van der Beek AJ: Determinants of physicians' communication behaviour in disability assessments. *Disabil and Rehab* 2011, **33**:1157-1168.
15. Schellart AJM, Zwerver F, Knol DL, Anema JR, Van der Beek AJ: Development and reliability of performance indicators for measuring adherence to a guideline for depression by insurance physicians. *Disabil Rehabil* 2011, **33**:2535-43.
16. Hamilton M: A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960, **23**:56-62.
17. Twisk JWR: *Applied longitudinal data analysis for epidemiology*. A practical guide. Cambridge University Press, 2003.
18. Adams AS, Soumeray SB, Lomas J, Ross-Degnan D: Evidence of self-report bias in assessing adherence to guidelines. *Int J Qual Health Care* 1999, **11**:187-92.
19. Millstein SG: Utility of the theories of reasoned action and planned behavior for predicting physician behavior: a prospective analysis. *Health Psychol* 1996, **15**:398-402.
20. Grimshaw JM, Eccles MP, Steen N, Johnston M, Pitts NB, Glidewell L, et al.: Applying psychological theories to evidence-based clinical practice: Identifying factors predictive of lumbar spine x-ray for low back pain in UK primary care practice. *Implement Sci* 2011, **6**:55.
21. Steenbeek R, Schellart AJ, Mulders H, Anema JR, Kroneman H, Besseling J: The development of instruments to measure the work disability assessment behaviour of insurance physicians. *BMC Public Health* 2011, **11**:1.
22. Karbach U, Schubert I, Hagemeister J, Ernstmann N, Pfaff H, Höpp HW: Physicians' knowledge of and compliance with guidelines. An exploratory study in cardiovascular diseases. *Dtsch Arztebl Int* 2011, **108**:61-69.
23. Bonetti D, Johnston M, Pitts NB, Deery C, Ricketts I, Tilley C, et al.: Knowledge may not be the best target for strategies to influence evidence-based practice: using psychological models to understand RCT effects. *Int J Behav Med* 2009, **16**:287-93.
24. Coomarasamy A, Khan KS: What is the evidence that postgraduate teaching in evidence based medicine changes anything? A systematic review. *BMJ* 2004, **329**:1017.
25. Heselmans A, Donceel P, Aertgeerts B, Van de Velde S, Ramaekers D: The attitude of Flemish occupational health physicians towards evidence-based occupational health and clinical practice guidelines. *Int Arch Occup Environ Health* 2010, **83**:201-8.



## Appendix 6.1: ASE determinants questionnaire.

### Attitude concerning the use of the guidelines for depression

All items are scored with Likert Scale (1-5) 1 = strongly disagree, 5 = strongly agree, if not reported otherwise

1. The guidelines for depression can support the IP with making complicated decisions
2. The guidelines for depression stimulates professionalization of the IPs
3. Working in concordance with the guidelines for depression is too rigid for the individual client
4. The guidelines for depression provide for an increase in quality of assessment
5. The guidelines for depression can improve the relationship between IP and the client
6. The guidelines for depression are a threat to the autonomy of the IP
7. Working in concordance with the guidelines for depression hinders professionals in making them familiar with new insights concerning depression
8. I agree with the content of the guidelines for depression
9. My attitude towards the guidelines for depression is positive

### Self-efficacy concerning the use of the guidelines for depression

1. I feel sufficiently equipped for applying the guidelines for depression
2. The guidelines for depression positively influence the quality of my assessments in practice
3. The guidelines for depression to me are useful for:
  - A. Getting my assessment structured
  - B. Taking away my doubts
  - C. Strengthening my process of taking decisions
  - D. Writing down my work disability report
  - E. Preparing my assessment interview
  - F. Freshening up my knowledge
4. The information presented in the guidelines for depression to me is: Too complex; Just right; Too simple, or Not known
5. How do you think of the clarity of the following aspects of the guidelines for depression?
  - A. The aim of the guidelines, to me is: Not clear; A little bit clear; or Completely clear
  - B. Assessing the prognosis to me is: Not clear; A little bit clear; or Completely clear

### Knowledge and Skills concerning the use of the guidelines for depression

1. I have sufficient knowledge to apply the guidelines for depression
2. I have the skills to work in concordance with the guidelines for depression
3. I feel needs for further training and exercising in the use of the guidelines for depression
4. I am able to organize my work in order to apply the guidelines for depression
5. Learning to work in concordance with the guidelines for depression takes more time from me, than I have at disposal
6. I have difficulties to integrate the use of the guidelines for depression in my daily work routine
7. Present (disability) legislation leaves enough room for working in concordance with the guidelines for depression
8. I believe that applying the guidelines for depression is practically feasible

**Intention to use the guidelines for depression**

1. I have the intention to use or keep using elements of the guidelines for depression
2. I expect to use elements from the guidelines for depression in the near future
3. I am intending to use or keep using the complete guidelines for depression
4. I think the guidelines for depression are useful for taking decisions concerning the assessment of the work limitations
5. Working in concordance with the guidelines for depression should be compulsory
6. The clients, which I assess, benefit from the implementation of the guidelines for depression
7. To my opinion, there are clients, for whom the guidelines for depression are not applicable
8. The guidelines for depression probably will be used in appeal cases
9. The guidelines for depression probably will contribute to a decrease in lost appeal cases
10. The guidelines for depression will contribute to a higher uniformity in the work disability assessments of clients with depression

# **Chapter**

---

# **7**

## **The influence of applying insurance medicine guidelines for depression on disability assessments**

**AJM Schellart  
F Zwerver  
JR Anema  
AJ van der Beek**

## Abstract

**Introduction:** In the current study we report on the effects of an implementation strategy in the form of a training programme on the assessed work limitations of a client with depression by insurance physicians (IPs) participating in an RCT. These assessed work limitations of a client were in the form of scores on the List of Functional Abilities (LFA).

**Method:** We conducted a randomised controlled trial (RCT) for IPs in which we compared the intervention of a specially developed training programme to the usual methods of implementation and training currently used. The outcome was the mean sum score and the inter-rater reliability (Intraclass Correlation Coefficient, ICC) of the LFA scores. These LFA scores were obtained from the IPs participating in the RCT for the work limitations of the cases presented in different videos, two videos before the training and two after the training of the intervention group.

**Results:** At baseline, the intervention group (IG) consisted of 21 IPs and the control group (CG) of 19. For one participant of the IG and for one of the CG the LFAs the two case reports after training were not available. Before training the sum scores for the first case report did not differ significantly between the groups, while the mean sum score was higher in the IG than in the CG for the second case report. For both case reports after training a higher score was found in the IG than in the CG. The inter-rater reliability measured for the two case reports before training was about the same in the IG and the CG: 0.64 and 0.65, respectively. For the two case reports after training, the ICC was higher in the IG than in the CG: 0.69 and 0.54, respectively. This difference was not statistically significant however.

**Conclusion:** It would appear that the implementation of a specially designed training programme on guidelines for depression may lead to greater inter-rater reliability in the assessments by insurance physicians of the work limitations of clients with depression. It is, however, important to note that insurance physicians who receive training may find more work limitations than those who do not.



## Introduction

We have previously investigated whether an implementation strategy that meets the needs of insurance physicians (IPs) leads to better adherence to guidelines than the usual implementation employed by the Dutch National Institute for Employee Benefits Schemes [1]. To this end we have developed a training programme using interventions that teach IPs how to apply the insurance medicine guidelines for depression [2] when performing assessments for work limitations. The efficacy of this implementation strategy was investigated in a randomised controlled trial (RCT), in which a group of IPs trained in applying the guidelines for depression were compared with a control group. We have demonstrated that IPs trained in applying the guidelines for depression scored significantly higher on guideline adherence and on knowledge of the guidelines for depression than IPs in the control group [3].

In the current study we report on the effects of this implementation strategy in the form of a training programme on the work limitations of a client with depression by insurance physicians participating in the RCT. Our aim was to study this effect in the form of scores on the List of Functional Abilities or LFA [4], which represent a combination of the number as well as the severity of work limitations. The LFA is partly based on the International Classification of Functioning (ICF) [5]. The ICF has internationally been used for qualifying the level of functioning in disability assessments [6, 7]. The following questions were therefore central to our research:

- I What is the influence of the training programme on the work limitations?
- II What is the influence of the training programme on the inter-rater reliability between the LFA scores of the participating IPs?

Previous research by Spanjer et al. [8] has shown that the more information about the client the IP has, the higher the number of work limitations the IP will find. A study by Schellart et al. [9] of inter-doctor variation between assessments by IPs found that greater adherence to the rules by IPs leads to a greater number of clients being assessed as the highest category of work disability. Based on these studies, our thoughts in the current study are that our intervention – a specific training programme on applying the guidelines for depression – will possibly lead to a more systematic overview of disorders and therefore to the finding of a higher number of work limitations in the intervention group than in the control group. We also think that our training programme may cause IPs to assess work limitations in a more uniform manner based on the information available. If this is indeed the case then the inter-rater reliability of the completed LFAs based on the same case reports should be greater in the intervention group than in the control group. Based on these thoughts we formulated the following hypotheses:

- 1) Training in guidelines for depression will result in more work limitations, because adherence to the guidelines leads to a more complete overview of disorders and the resulting work limitations, based on the information available.

2) Training in guidelines for depression will result in higher inter-rater reliability between IPs: after following the training programme the IPs will assess work limitations in a more uniform manner.

## Methods

### Design

To determine the efficacy of a specially developed strategy for implementation of the guidelines for depression [1], we conducted a randomised controlled trial (RCT) in which we compared an intervention group with a control group. In this RCT we compared the intervention of a specially developed training programme with the usual methods of implementation and training currently in use by the social security agency.

The intervention was a training programme designed for IPs, in which they learned to apply the guidelines for depression [2]. This programme, together with baseline and follow-up measurements, was integrated into a four-day postgraduate course located at the Netherlands School of Public and Occupational Health (NSPOH).

While the intervention group was trained in applying the guidelines for depression, the control group received an alternative programme of training in motivational interviewing that did not conflict with the intervention programme. The RCT took three days within a period of two weeks in March 2009. After the RCT ended, the control group received the same training as the intervention group, while the intervention group received the alternative programme. This was planned as the fourth day of the course, which was held three months later at the end of June 2009.

By using actors simulating four different case reports on video, we managed to create a laboratory setting in which we could measure the work disability assessments of clients with depression by each IP. In these videos the role of the client was played by four different actors, while the role of the IP was played by two 'real' IPs, independently selected for this purpose. The training programme was designed to be also applied in practice. The Ethics Committee of the VU University Medical Centre granted approval for the study design and the RCT was accepted by the Netherlands Trial Register under number NTR1863.

### Participants

In January 2009, IPs employed by the social security agency were invited to take part in a postgraduate course in applying the guidelines for depression, given in the period from March to July 2009. The inclusion criteria were that individuals should be registered as insurance physicians, or still in training as such, and should be conducting disability assessments of clients as commissioned by the Institute. The NSPOH was responsible for enrolment of participants, who also provided written informed consent to take part in the study.



The participants were allocated in order of registration to either the intervention group or the control group by using a random-sequence table. Participants who were not available on the planned dates were excluded from the trial. The participants were informed about the fact that the course was part of a research project, but they were not informed about the design of the entire project, i.e. the various measurements and the group they participated in.

### Data collection

Data were collected at the NSPOH during the period of the training course. At baseline (pre-intervention) and at follow-up (post-intervention) each IP assessed the work limitations of two clients, played by actors, who were presented separately on video. The actors played clients with depression, reconstructed from real case reports. The actors played their roles on the basis of extensive scripts, with room for improvisation. The videos showed the disability assessment encounter between a client (actor) and an independent IP (not a participant in the RCT), who had been briefed to perform the assessment in complete accordance with the guidelines for depression. The decision phase of the assessment encounter was not shown on the video. The participating IPs completed their medical disability reports, including the LFA, immediately after watching each client on the video. All reports and completed LFAs were collected directly afterwards. The researchers were blinded for the collection of data and an independent research assistant coded the data.

### Outcomes

The primary outcome of the RCT was guideline adherence, measured using performance indicators. A detailed description of the development and reliability of these performance indicators has been published elsewhere [10], as has the effect of the intervention on guideline adherence [3].

Secondary outcome in the RCT was LFA scores. These LFA scores were assessed by the IP participating in the RCT for the work limitations of the clients presented in the four videos. The LFA consists of six sections containing a total of 106 items: I personal functioning (30 items), II social functioning (17 items), III adjusting to the physical environment (13 items), IV dynamic movements (31 items), V static posture (11 items), and VI working hours (4 items). A large-scale study (of 51,000 disability assessments) into the dimensions behind these items [11] discovered 16 dimensions, each forming an scale. The internal reliability of the scales (Cronbach's alpha) was generally acceptable (alpha 0.60-0.75) to good (alpha >0.75) or even very good (alpha >0.85). Only one dimension – communication – had an unacceptable level of internal reliability (alpha 0.53). In a follow-up study using a second order factor analysis [12], 14 of these 16 scales (excluding communication and working hours) were further reduced to four scales:

- 1) mental abilities: limitations in coping with various mental task demands

- 2) general physical abilities: limitations covering various aspects of the musculoskeletal system
- 3) autonomy: limitations in being able to act autonomously in the working situation
- 4) manual skills and grip strength limitations.

Since the internal reliability of this last scale was very low (alpha 0.46), items on this scale were included in the scale for general physical ability, a possibility demonstrated by another study of LFA data from 84,000 disability assessments [9]. The three scales in the mentioned study had an acceptable level of reliability (alphas were 0.69 for scale 1, 0.72 for scale 2, and 0.75 for scale 3 including manual skills and grip strength). Hence, in the current study we used these three scales, with an additional separate scale for working hours, which had a very good internal reliability (alpha 0.97) [11].

### Analyses

To address the first hypothesis, we used an unpaired t-test to analyse differences in the mean sum scores of the four scales between the intervention group and the control group for each case report (four case reports: the first two pre-intervention, the other two post-intervention). To examine whether correction was necessary for the influence of any unequal distribution of background variables between the intervention group and the control group, we performed regression analysis using the relevant background variable as covariate.

To address the second hypothesis regarding inter-rater reliability, we performed analyses using linear mixed models, which enable modelling of variances (and covariances) and provide the possibility of accounting for hierarchical data [13]. We used the variances to calculate the intraclass correlation coefficient (ICC, with values ranging between 0 and 1) [14]. A higher ICC is an indication of greater degree of inter-rater reliability. We also calculated whether the difference between the ICCs of the intervention group and the control group was significantly different from zero. For a more detailed description of the statistical analysis we refer to Appendix 7.1. All analyses were performed using SPSS 15.0 [13].

## Results

### Participants

Between January and March 2009 a total of 43 insurance physicians applied to take part in the course. At the time of the RCT all participating IPs were actively conducting disability assessments. Twenty-one IPs were allocated to the control group and 22 to the intervention group. One of the IPs who was allocated to the intervention group withdrew from the course and 2 IPs who were originally allocated to the control group were not available on the planned dates. All three were excluded from the RCT. At



baseline, therefore, the control group (CG) consisted of 19 IPs and the intervention group (IG) of 21. For one CG participant and for one IG participant the LFAs of the two case reports after training were not available.

The separate baseline characteristics were equally distributed across both groups, apart from one variable (see Table 7.1). Although the mean number of clients with depression assessed by an IP per month was significantly higher in the CG, regression analysis demonstrated that this variable had no major effect on the magnitude of the sum scores of the four scales in the CG and IG for the four separate case reports. Correction for this variable in the analyses was therefore not necessary.

**Table 7.1:** Baseline characteristics of insurance physicians in control group (CG) and intervention group (IG).

Baseline characteristics	CG (n = 19)	IG (n = 21)
	Mean (sd) or percentage	
Age in years	50.5 (6.7)	51.1 (6.2)
Male	47%	52%
Weekly working hours	31.8 (9.9)	31.1 (9.2)
Years working as physician	21.7 (6.4)	23.5 (5.1)
Registered as insurance physician	84%	86%
Years working as insurance physician	15.4 (8.1)	15.6 (7.9)
Number of clients with depression assessed per month*	9.3 (5.6)	5.3 (3.7)
Assessment time for depressed clients (minutes)	136.3 (62.3)	153.7 (48.4)
Assessments under the new disability act	68%	52%
Employee of the Institute	79%	81%

\* Significant difference between both groups ( $p < 0.05$ ); Institute: the Dutch Institute for Employee Benefits Schemes.

### Outcomes

Table 7.2 shows the mean scale scores (with standard deviation) for each LFA scale and the corresponding sum scores of the scales for the first two case reports before training, for both the control group (CG) and the intervention group (IG). For case report 1, most participants filled in items on the scales for working hours and mental abilities. About half the participants filled in items on the scale for physical abilities. Hardly any items on the scale for autonomy were filled in. The means of the sum score did not differ significantly over the four scales between CG and IG ( $p = 0.229$ ). For case report 2, again most participants filled in items on the scale for mental abilities. This was also mainly

the case for the scale for working hours in the IG, but not in the CG: in the CG about half the participants filled in work limitations on this scale. For the scales for autonomy and physical abilities, participants in the IG filled in items about twice as often as those in the CG. In the IG the mean sum score over the four scales was significantly higher than this mean sum score in the CG ( $p = 0.013$ ).

**Table 7.2:** Mean scale scores (sd) of LFA scales for two case reports before training\*.

	working hours	autonomy	physical abilities	mental abilities	sum score
<b>Case report 1: CG</b>					
N (n)	19 (17)	19 (2)	19 (8)	19 (19)	19 (19)
Mean (sd)	3.68 (2.08)	0.11 (0.32)	1.53 (2.27)	6.74 (3.18)	12.05 (5.10)
<b>Case report 1: IG</b>					
N (n)	21 (18)	21 (1)	21 (10)	21 (20)	21 (20)
Mean (sd)	4.48 (2.27)	0.10 (0.44)	1.95 (3.11)	8.00 (3.96)	14.52 (7.34)
<b>Case report 2: CG</b>					
N (n)	19 (10)	19 (5)	19 (2)	19 (19)	19 (19)
Mean (sd)	2.32 (2.69)	0.68 (1.45)	0.32 (1.16)	9.00 (3.80)	12.32 (4.80)
<b>Case report 2: IG</b>					
N (n)	21 (18)	21 (11)	21 (6)	21 (21)	21 (21)
Mean (sd)	3.95 (2.42)	0.81 (1.03)	1.48 (3.33)	11.24 (4.07)	17.48 (7.37)

\* LFA = List of Functional Abilities; N = number of insurance physicians; n = number of insurance physicians who filled in disabilities for (the scale of) the LFA; sd = standard deviation. The difference of the mean sum scores over the four LFA scales between control group (CG) and intervention group (IG) is not significant for case report 1 ( $p=0.229$ ), but is significant for case report 2 ( $p=0.013$ ).

For the two case reports after intervention (case reports 3 and 4, see Table 7.3), the mean sum scores in the IG were significantly higher than those in the CG ( $p = 0.023$ ). For case report 3 few participants filled in items on the scales for autonomy and physical abilities. For case report 4 the tendencies and distribution of the CG and IG were only of interest for the scale for mental abilities. Here again the mean sum score in the IG was significantly higher than that in the CG ( $p = 0.04$ ).



**Table 7.3:** Mean scale scores and sum scores of LFA scales for two case reports after training\*.

	working hours	autonomy	physical abilities	mental abilities	sum score
<b>Case report 3: CG</b>					
N (n)	18 (6)	18 (3)	18 (0)	18 (16)	18 (16)
Mean (sd)	1.11 (2.27)	0.39 (0.98)	0.00 (0.00)	4.94 (3.81)	6.44 (6.25)
<b>Case report 3: IG</b>					
N (n)	20 (16)	20 (3)	20 (4)	20 (19)	20 (20)
Mean (sd)	3.80 (2.33)	0.30 (0.80)	1.25 (3.02)	8.70 (3.80)	14.05 (6.44)
<b>Case report 4: CG</b>					
N (n)	18 (0)	18 (0)	18 (0)	18 (16)	18 (16)
Mean (sd)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	4.00 (2.38)	4.00 (2.38)
<b>Case report 4: IG</b>					
N (nu)	20 (5)	20 (0)	20 (0)	20 (20)	20 (20)
Mean (sd)	0.60 (1.14)	0.00 (0.00)	0.00 (0.00)	5.45 (2.42)	6.05 (2.87)

\*LFA = List of Functional Abilities; N = number of insurance physicians; n = number of insurance physicians who filled in disabilities for (that scale of) the LFA; sd = standard deviation. The difference of the mean sum scores over the four LFA scales between control group (CG) and intervention group (IG) is significant for both case report 3 ( $p=0.001$ ) and case report 4 ( $p=0.023$ ).

Table 7.4 shows the results of the mixed models analysis (parameters and standard errors), and of the ICC calculation for the presented case reports before training (case reports 1 and 2) and after training (case reports 3 and 4). For the case reports before training (case reports 1 and 2) the ICCs were similar (0.65 for the CG and 0.64 for the IG). For the case reports after training (case reports 3 and 4) the ICC in the IG was 0.69 and the ICC in the CG was 0.54. Upon testing, however, both differences in the ICCs between the IG and the CG were not statistically significantly different from zero. The difference in ICC between the IG and CG (95% confidence interval) was: -0.01 (-0.56; 0.54) for case reports 1 en 2, and 0.14 (-0.35; 0.68) for case reports 3 and 4.

To determine whether the difference in ICCs between the CG and the IG as shown in Table 7.4 might have been influenced by the low number of observations for some of the abilities scales, the same analysis was conducted using either three scales – excluding the scale for physical abilities – or using two scales, i.e. using only the scales for working hours and mental abilities. Once more, the ICCs of the IG and the CG (see Table 7.5) were about the same for case reports 1 and 2 and higher for case reports 3 and 4 (0.21 higher when using 3 scales and 0.16 higher when using 2 scales). The differences in ICC between the IG and the CG were again in all cases not statistically significantly different from zero (not shown here).

**Table 7.4:** Results of the mixed models analysis and ICC calculation, with scores of four LFA scales\*.

	Case reports 1 and 2		Case reports 3 and 4	
	CG	IG	CG	IG
residual	5.28 (0.07)	6.37 (0.82)	2.52 (0.35)	4.06 (0.54)
case report	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
scale (case report)	10.29 (5.65)	14.64 (7.99)	3.89 (2.16)	9.98 (5.45)
respondent	0.32 (0.34)	1.61 (0.84)	0.43 (0.35)	0.28 (0.36)
case report * respondent	0.00 (0.00)	0.18 (0.60)	0.34 (0.34)	0.25 (0.43)
ICC	0.65	0.64	0.54	0.69
(95% confidence interval)	(0.33-0.84)	(0.32-0.83)	(0.21-0.76)	(0.37-0.86)

\*Estimated for the case reports before training (case reports 1 and 2) and after training (case reports 3 and 4) in control group (CG) and intervention group (IG), with linear mixed models (parameters, standard errors) and variance components for mixed models (ICCs and 95% confidence interval); the four disability scales are: working hours, autonomy, physical abilities, and mental abilities; ICC = Intraclass Correlation Coefficient.

**Table 7.5:** Results of the ICC calculation, with scores of three and two LFA scales respectively\*.

	Case reports 1 and 2		Case reports 3 and 4	
	CG	IG	CG	IG
ICC (3 scales)	0.65	0.71	0.51	0.72
ICC (2 scales)	0.49	0.51	0.46	0.62

\*Estimated for the case reports before training (case reports 1 and 2) and after training (case reports 3 and 4) in control group (CG) and intervention group (IG), with variance components for mixed models; the three disability scales are: working hours, autonomy, and mental abilities; the two disability scales are: working hours and mental abilities; ICC = Intraclass Correlation Coefficient.

## Discussion

### Main findings

The results of this study show that before training the sum scores for the first case report did not differ significantly between the groups, while for the second case report the mean sum score was significantly higher in the IG than in the CG. For the two case reports after training, we saw a significantly higher score in the IG than in the CG.

The inter-rater reliability measured for the two case reports before training and using four scales was about the same in the CG and the IG. For the two other case reports after training, the ICC was 0.69 for the IG and 0.54 for the CG. This difference was not statistically significant however.

### Interpretation and comparison with other studies

The training programme on applying the guidelines for depression resulted in more work limitations. For the same case report, IPs who received training filled in more work limitations in the LFA than the IPs who did not receive training. This difference is most noticeable in case report 3.

Post-intervention data showed that the group of IPs who were given training in applying the guidelines had a higher degree of consistency when filling in the LFA than the IPs in the control group. Apparently the implementation strategy contributed to more uniformity in work limitations assessments by IPs. This ties in well with earlier research into variation in work disability assessments [15, 16]. In terms of financial and social consequences, such variation is unwanted for both the client and society and in our opinion might be reduced by the use of standardised methods of assessment, as occurs when guidelines are applied. The fact that applying guidelines results in a more uniform judgment ties in well with the idea that reducing medical ambiguity or uncertainty also reduces variation between doctors [17, 18].

It is striking that the differences between the two groups with regard to the scale for working hours were considerable (except for case report 1), both before and after training. Working hours limitation is a strong determinant for the end result of the assessment: the degree of work disability assigned to the client. Another study into variations in disability assessments also found little consistency between IPs regarding the work limitation scale for working hours [16]. The scale for working hours even has its own guidelines, separate from those specific to diagnosis [19].

Our results confirm the trends posed in the two hypotheses. We have shown that IPs trained in using the guidelines apply more work limitations than untrained IPs. In another study of ability assessments of clients with depression, the use of a work ability checklist actually led to findings of higher levels of work ability, without a reduction in the variation of assessment results [20]. One possible explanation for this is that the emphasis in the aforementioned study was on work ability rather than on work limitations as in the depression guidelines. Incidentally, the ICC in that study was of the similar magnitude to that found in the current study's pre-intervention measurements, namely 0.64.

The training programme taught the IPs to conduct systematic and thoroughly justified disability assessments in accordance with the guidelines. Apparently this method of assessment leads to a higher number of work limitations than is usually the case. The reason for this might be that IPs who adhere more closely to guidelines interpret the information provided more strictly than usual. After all, the information concerning the client was provided by means of a case report on video, which was the same for all IPs. The IPs themselves were not able to ask the client any questions. Therefore, in the daily practice of IPs – where interviews form an influential part of a disability assessment – the difference between the groups may well be greater: the trained IP, actively applying the guidelines, will



make further enquiries of the client regarding aspects such as sleep disorders. The existence of sleep disorders may then in turn influence how the IP fills in the LFA.

### Strengths and weaknesses

This study has several strengths. Firstly, the active form of the four 'real life' case reports on video, which simulate the daily practice of an IP, is more effective than written case reports [21]. Secondly, the fact that the two case reports presented before the training programme were different to the two after training prevents any confounding learning effect that occurs when a case report is presented for the second time. Thirdly, the suitability of the four scales drawn up on the basis of the LFA scientific research has already been established by statistical analysis in previous studies [9, 11, 12]: the difference in the means has been tested using the sum scores of the four scales, which are a valid measure of the number and severity of the limitations, since they are not influenced by the distribution over the four scales. Finally, to determine inter-rater reliability, an empirically tested method was used to calculate the ICCs (see appendix): the differences between the ICCs of the CG and IG were statistically tested for their difference from zero.

The study also has a number of weaknesses. To start with, it may be difficult for IPs to complete an LFA based purely on a video, a factor that was not looked at in this study. Another weakness is the question of what to do about items marked as 'no limitations found': should this be considered as missing data, or as an actual assessment of there being no limitations, or at least no severe limitations? We attempted to accommodate this weakness by also analysing inter-rater reliability while excluding the scales that had only a few observations. A further weakness is the fact that the pre-intervention data already showed a significant difference in the severity and number of limitations between the intervention group and the control group. Finally, since the case reports presented before and after the training programme were not necessarily comparable, the ICCs from before and after training were not comparable within each group (CG and IG). It was, therefore, not possible in the IG to test whether there was an increase in inter-rater reliability after the training programme.

### Practical relevance

The findings of this study provide a point of consideration for insurance medicine. IPs should be aware of the fact that collecting information about a client in a structural manner, as when following a guideline, can lead to the finding of more work limitations in that client. The IP should not lose sight of the importance of work participation and should focus on the work ability of the client. In addition, it would appear that IPs have difficulty reaching uniformity in applying the 'reduced working hours' standard [20]. We recommend a separate training programme for IPs to teach them to apply this standard, preferably according to the existing disease-specific guidelines.

Policy makers should be aware that although it is possible to improve the inter-rater reliability between IPs for disability assessments, there is still space for professional autonomy and variation in assessments, even after guidelines have been implemented. IPs cannot be completely constrained to a guideline and a guideline cannot be fully comprehensive to cover all possible situations. This study found a maximum ICC of 0.69, and not of 1.00. Since disability assessments are, and will remain, human activities, a certain degree of variation within professional guidelines is acceptable.

### Conclusion

There are indications that the implementation of a specially designed training programme on guidelines for depression may lead to greater inter-rater reliability in the assessments by insurance physicians of the work limitations of clients with depression. It is, however, important to note that insurance physicians who receive training may find more work limitations than those who do not. Whether this possible rise in work limitations found might also lead to a higher degree of work disability requires further investigation.

### Acknowledgements

The authors wish to thank the IPs who participated in this research. The Research Center for Insurance Medicine AMC-UMCG-UWV-VU University Medical Center, in Amsterdam, is a joint initiative of the Academic Medical Center (AMC), the University Medical Center in Groningen (UMCG), the Dutch Institute for Employee Benefits Schemes (UWV), and the VU University Medical Center (VUMC). This trial was funded by the Dutch Institute for Employee Benefits Schemes and the Netherlands' School for Public and Occupational Health. FZ, JRA, and AJMS are (partially) funded by UWV. The study sponsor had no role in the study design, in the collection, analysis or interpretation of the data, in the writing of the case reports, or in the decision to submit the paper for publication. The design of this study was laboratorial and for data collection fictitious but realistic case-reports were used. Consequently, the Medical Ethics Committee agreed with the design. The full trial protocol can be accessed at the webaddress of the Netherlands' Trial Register (NTR): <http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=1863>.

### Authors' contributions

The authors declare that they participated in the study and made the following contributions to the study. AJMS, FZ, JRA and AJvdB contributed to the conception and design of this study. AJMS and FZ contributed to the analysis. AJMS and FZ wrote the manuscript. JRA and AJvdB revised and commented on the manuscript. AJMS and AJvdB will act as guarantors of this study. AJvdB had full access to all data in the study and

had final responsibility for the decision to submit for publication. All authors read and approved the final manuscript.

### Competing interests

The authors declare that they have no competing interests.

### References

- Zwerver F, Schellart AJM, Anema JR, Rammeloo K, Van der Beek AJ: Intervention mapping for the development of a strategy to implement the insurance medicine guidelines for depression. *BMC Public Health* 2011, **11**:9.
- Health Council: *Insurance Medicine Guidelines for Depression*. Den Haag: Gezondheidsraad, 2006.
- Zwerver F, Schellart AJ, Knol DL, Anema JR, Van der Beek AJ: An implementation strategy to improve the guideline adherence of insurance physicians: an experiment in a controlled setting. *Implement Sci* 2011, **6**:131.
- Lisv: *ClaimBeoordelings- en BorgingsSysteem (CBBS) (List of Functional Abilities)*. Amsterdam: Lisv, 2002.
- World Health Organization: *International classification of functioning, disability and health*. Geneva: WHO; 2002.
- Brage S, Donceel P, Falez F, Working Group of the European Union of Medicine in Assurance and Social Security: Development of ICF core set for disability evaluation in social security. *Disabil Rehabil* 2008, **30**:1392-6.
- Østerås N, Brage S, Garratt A, Benth JS, Natvig B, Gulbrandsen P: Functional ability in a population: normative survey data and reliability for the ICF based Norwegian Function Assessment Scale. *BMC Public Health* 2007, **7**:278.
- Spanjer J, Krol B, Popping R, Groothoff JW, Brouwer S: Disability assessment interview: the role of detailed information on functioning in addition to medical history-taking. *J Rehabil Med* 2009, **41**:267-72.
- Schellart AJ, Mulders H, Steenbeek R, Anema JR, Kroneman H, Besseling J: Inter-doctor variations in the assessment of functional incapacities by insurance physicians. *BMC Public Health* 2011, **11**:864.
- Schellart AJ, Zwerver F, Knol DL, Anema JR, Van der Beek AJ: Development and reliability of performance indicators for measuring adherence to a guideline for depression by insurance physicians. *Disabil Rehabil* 2011, **33**:2535-43.
- Broersen JP, Mulders HP, Schellart AJ, Van der Beek AJ: The dimensional structure of the functional abilities in cases of long-term sickness absence. *BMC Public Health* 2011, **11**:99.
- Broersen JP, Mulders HP, Schellart AJ, Van der Beek AJ: The identification of job opportunities for severely disabled sick-listed employees. *BMC Public Health* 2012, **12**:156.
- SPSS 15.0 Command Syntax Reference 2006. Chicago III: SPSS Inc; 2006.
- Molenberghs G, Laenen A, Vangeneugden T: Estimating reliability and generalizability from hierarchical biomedical data. *J Biopharm Stat* 2007, **17**:595-627.
- Spanjer J, Krol B, Brouwer S, Groothoff JW: Sources of variation in work disability assessment. *Work* 2010, **37**:405-11.



- Spanjer J, Krol B, Brouwer S, Groothoff JW: Inter-rater reliability in disability assessment based on a semi-structured interview report. *Disabil Rehabil* 2008, **30**:1885-90.
- Wennberg JE, Barnes BA, Zubkoff M: Professional uncertainty and the problem of supplier-induced demand. *Soc Sci Med* 1982, **16**:811-824.
- Eisenberg JM: Physician utilization: the state of research about physicians' practice patterns. *Med Care* 2002, **40**:1016-35.
- Lisv: *Standaard verminderde arbeidsduur (Reduced working hours standard)*. Amsterdam: Lisv; 2007.
- Slebus FG, Kuijer PP, Willems JH, Frings-Dresen MH, Sluiter JK: Work ability assessment in prolonged depressive illness. *Occup Med* 2010, **60**:307-9.
- Berkhof M, Van Rijssen HJ, Schellart AJ, Anema JR, Van der Beek AJ: Effective training strategies for teaching communication skills to physicians: an overview of systematic reviews. *Patient Educ Couns* 2011, **84**:152-62.

### Appendix 7.1: Statistical method for calculating intraclass correlation coefficients.

The intraclass correlation coefficients (ICCs) were calculated separately for the intervention group and the control group, both for the two case reports before the intervention and for the two case reports after the intervention. For this calculation we used a formula derived from the so-called generalizability theory [A1, A2], with the following variance components of a linear mixed model [A3, A4]: the two case reports, the four LFA scales nested within the case reports, the IPs, the interaction between the case reports and the IPs, and the residual variance. For the ICC calculation, the sum of the variance components for the case reports and the scales within the case reports formed the universe score, while the sum of the other variance components formed the absolute error variance. The ICC is defined as the ratio of the universe score to the sum of the universe score and the absolute error variance. The 95% confidence intervals were calculated from the variance components for a mixed model [A3] using Fisher's Z transformation and the delta method [A5]. These methods were also used to calculate whether the difference between the ICCs of the intervention group and the control group was significantly different from zero (for a 95% confidence interval).

### References in Appendix

- Shavelson RJ, Webb NM, Rowley GL: Generalizability theory. *American Psychologist* 1989, **44**:922-932.
- Shavelson RJ, Webb NM: *MMSS generalizability theory: a primer*. Newbury Park, CA: Sage 1991.
- SPSS 15.0 Command Syntax Reference 2006. Chicago III: SPSS Inc; 2006.
- Molenberghs G, Laenen A, Vangeneugden T: Estimating reliability and generalizability from hierarchical biomedical data. *J Biopharm Stat* 2007, **17**:595-627.
- Euser AM, Le Cessie S, Finken MJ, Wit JM, Dekker FW: Reliability studies can be designed more efficiently by using variance components estimates from different sources. *J Clin Epidemiol* 2007, **60**:1010-1014.

# Chapter

---

# 8

## **Implementation of a strategy for the insurance medicine guidelines for depression: a process evaluation**

**F Zwerver  
K Groenewoud  
AJM Schellart  
JR Anema  
AJ van der Beek**

## Abstract

**Background:** We developed an implementation strategy for the insurance medicine guidelines for depression and carried this out via a post-graduate course for insurance physicians. Learning objective for the physicians in the training was to apply the guidelines for depression. In this study we evaluate their experiences of the implementation strategy.

**Methods:** Insurance physicians conducting disability assessments were invited to attend a post-graduate course on the implementation strategy for the guidelines of depression in which a controlled experiment was embedded. Data were collected from the participating insurance physicians using questionnaires applied directly after the intervention and at three-month follow-up.

**Results:** Of the 797 insurance physicians invited, 43 participated. Reach amounted to 5%. The response to the questionnaires was 98%. The participants appraised the implementation strategy with a total score of 7.7 out of 10, and 81% expressed an expectation of improving their assessments of clients with depression. After three months their satisfaction with the implementation strategy was sustained. Changing work routines and the time needed for applying the guidelines were perceived as barriers to use the implementation strategy.

**Conclusions:** The reach of the newly developed implementation strategy for the guidelines for depression was poor: only 42 IPs out of 797 invited attended the post-graduate course. However, the results show that the implementation strategy worked well in a controlled setting for all participants. They were satisfied with the training and tools both immediately after the training and after a period of three months.



## Background

Worldwide, depression contributes increasingly to work disability [1-3]. Assessing disability of workers with depression is and will be a common and challenging task for insurance physicians (IPs). Guidelines should support physicians by providing them with evidence-based medicine (EBM) and with recommendations on how to use EBM in practice. However, many barriers to the implementation of guidelines might occur. Barriers to implementation are possible at different levels, such as the level of the individual behaviour of the physician, the organizational/contextual level, the level of the guidelines themselves and the implementation strategy level [4]. Sometimes guidelines can be hard to translate into practice and sometimes a patient (or client) seems not to fit into a specific guideline. In the case of depression, assessing the diagnosis might be difficult, because symptoms of depression can occur in many different ways or may be hidden. And how should a physician deal with the frequently occurring phenomenon of co-morbidity? Furthermore, once the diagnosis has been assessed, judgement of the severity of the disorder and the assessment of the client's functional limitations can easily lead to a wide inter-rater variability between physicians [5]. These few examples illustrate how difficult it can be to get EBM implemented properly. According to the guidelines, IPs are expected to conduct disability assessments uniformly, transparently, and based on evidence [6, 7]. Therefore, an effective implementation strategy for insurance medicine guidelines is needed in which attention is paid to translating the guidelines into the practice of IPs. We developed a new implementation strategy for the insurance medicine guidelines for depression and subsequently evaluated this strategy in a controlled experiment [8]. The strategy consisted of training and facilitating IPs with tools for learning to apply the guidelines. However, the implementation of guidelines for physicians has been described in literature as a complex and difficult process, influenced by many factors and often with disappointing results [4, 9, 10]. Therefore, knowledge of the barriers to and the facilitators of this implementation strategy is needed to evaluate and to improve the process of implementation.

This study aims to evaluate the process of the newly developed implementation strategy ie the training and tools for applying the guidelines for depression. Such a process evaluation carried out alongside an experiment could nuance the interpretation of the quantitative results and clarify the success and failure of an intervention [11, 12]. In this process evaluation, we could examine only the characteristics of the IPs as users and of the researchers as observers and not other elements such as the organization where the IPs' work. The aims of this study were to describe: (1) the recruitment and reach of the implementation strategy, (2) the dose delivered and dose received, to evaluate whether the implementation strategy was carried out as planned and how it was received, (3) the satisfaction and expectations of the IPs with the strategy, and (4) the perceived barriers to the implementation strategy.



## Methods

### Participants and design

This process evaluation was carried out alongside an experiment in a controlled setting on the efficacy of a newly developed implementation strategy for insurance physicians to assess and report on a client's work disability in accordance with the guidelines for depression. The Medical Ethics Committee of the VU University Medical Centre approved the study design. IPs, who conducted work disability assessments of clients on commission by the Dutch Institute for Employee Benefit Schemes (Institute), were invited to attend a post-graduate course in which they learned to apply the guidelines for depression. Inclusion criteria were: being registered as an IP or following the colloquium for registration as an IP, and conducting disability assessments at the Institute. The recruited IPs participated on voluntary basis.

### Intervention

The overall aims of the newly developed implementation strategy were to increase the guideline adherence and to improve the disability assessments and reports of IPs. The implementation strategy was planned as a one-day training course in the guidelines for depression. The training day consisted of various components. At the start of the day, a psychiatrist explained a number of important aspects of the assessment of depression supported by an attractive recent case of an employee with an atypical presentation of depressive symptoms. Focus points in this presentation were diagnostics, distinction between behaviour and disease, symptoms, the relationships between symptoms and disabilities, estimating the severity of depression, including use of the Hamilton Rating Scale for Depression (HRSD) [13], treatment, progression of the condition and co-morbidity. Next, the IP trainer introduced the evidence-based tools containing an easy to follow summary card with the most essential information on the guidelines, i.e. diagnostics (DSM-IV), co-morbidity, therapy and treatment and the ICF-model [14], a checklist of the main items of the guidelines, a checklist of disabilities connected to the severity of the depression based on a Delphi study [15] and the HRSD. Thereafter, the participants were separated into subgroups, each of them focussing on one of the main items of the guidelines. Each subgroup had to make an assessment of a client with depression presented on video, with the help of the toolbox. The form of the training was characterized by self-activation, interactivity and feedback. Subsequently, the IP trainer explained how to write down the findings and the conclusions of the assessment by using the essential elements of reasoning. The participants were given feedback on their performances by the IP trainer. A research assistant was present on all days of the post-graduate course and not only took care of the logistics but also made observational reports on the execution of the planned actions. More detailed information about the intervention can be found elsewhere [8].

### Process measures and data collection

Based on the Linnan and Steckler framework for process evaluations in public health we addressed the following elements: recruitment, reach, dose delivered by the implementation providers, dose received by the participants and barriers to use the implementation strategy as perceived by the participants [16]. Fidelity and implementation could not be measured because the design of this study was an experiment in a controlled setting. The data for this process evaluation were collected from the participating IPs using questionnaires and checklists. Data at baseline and directly after the course were collected from all IPs. Furthermore, data at three-month follow-up were collected from a subgroup of IPs (n=23). In Table 8.1 the data collection and the process measures are presented.

### Recruitment and reach

Recruitment refers to the sources and the procedures used to attract potential participants and the number of initially recruited IPs; the most common reasons for non-participation were registered.

Reach was defined as the number of IPs participating in the post-graduate course compared with the number invited. The representativeness of the participating IPs was determined by comparing those to the total group of IPs conducting disability assessments at the Institute.

### Dose delivered and dose received

Dose delivered was described as the number of components of the implementation strategy provided to the participants by the IP-trainers. This takes into account the extent to which the training was carried out in accordance with the programme plan of the implementation strategy.

The number of participants actually attending the course (dose received) was recorded by the research assistant using the attendance list. Furthermore, one of the IP-trainers and the research assistant wrote an evaluation report directly after the training day describing the engagement of the participants with the programme.

**Table 8.1:** Process measures and data collection.

Outcomes	Definition	Data-collection
Recruitment	- Sources and procedures used to recruit potential participants	- Checklist
	- Number of initially recruited IPs	- Checklist
	- Reasons for non-participation	- Checklist
Reach	- Number and characteristics of participating IPs	- IP registration forms  - 11 items at baseline
	- Representativeness of the IPs	- Additional data from the Institute
Dose delivered	- Amount of components of the implementation strategy delivered	- Checklist
	- The extent to which training was provided as intended	- Written reports by the research assistant and trainer-IP
Dose received	- Number of IPs who actually attended the training	- Checklist
Participant satisfaction and expectations	- Overall satisfaction with the training day	- 1 item on 1-10 scale (very unsatisfied to very satisfied)
	- Satisfaction with the different components of the training programme	- 2 items per component on 1-5 Likert scale (very unsatisfied to very satisfied)
	- Expectations about improvement in disability assessments and reports with the use of the evidence-based tools	- 2 items on 1-5 Likert scale (no improvement to much improvement)
	- Expectations about usefulness of the evidence-based tools in different components of applying GD	- 6 items on 1-5 Likert scale (not useful to very useful)
	- Intention to use the different evidence-based tools	- 2 items per tool on 1-5 Likert scale (no intention to use to intention to use)
	- Usefulness of training after three months	- 1 item on 1-5 Likert scale (not useful to very useful)
	- Needs for further training after three months	- 3 items on 1-5 Likert scale (need training to not need training)
	- Satisfaction with the evidence-based tools after three months	- 7 items per tool on 1-5 Likert scale (very unsatisfied to very satisfied)
	- Satisfaction about GD with the evidence-based tools in daily practice after three months	- 2 items on 1-5 Likert scale (very unsatisfied to very satisfied)
Barriers	- Barriers in implementation GD directly after the training and at three-month follow-up	- 22 items on 1-5 Likert scale (no barrier perceived to perceived as barrier)

IPs=Insurance physicians; GD=Guideline Depression; Institute=Dutch Institute for Employee Benefits Schemes.



## Participant satisfaction and expectations

Levels of overall satisfaction with the training course and its different components and expectations about any improvement in skills resulting from it were requested in a questionnaire directly after the course. Expectations regarding usefulness of the evidence-based tools and the intention to use the different tools were based on the experiences of the IPs as assessed directly after the course. Experiences three months after the course were collected from a subgroup of the IPs. In this three-month follow-up questionnaire, usefulness of the course, needs for further training and satisfaction with the tools in the IP's daily practice were requested.

## Perceived barriers

The perceived barriers to work with the evidence-based tools were also requested immediately after the training day and at three-month follow-up. Questions were based on the validated questionnaire 'Barriers and facilitators assessment instrument' from the Centre for Quality of Care Research (WOK) [17]. The questionnaire included 22 items concerning the professional (IP) and the innovation (evidence-based tools).

## Data analysis

The data were analyzed using descriptive statistics such as frequencies and means. SPSS 15.0 and Excel 2003 were used for the descriptive and statistical analyses. Cronbach's alphas were calculated for scales of items that were used for the outcome measures. Differences between measurements directly after training and measurements at three-month follow-up were calculated with a paired-sample T-test.

## Results

### Recruitment and reach

Between May 2008 and December 2008 IPs were invited to attend a post-graduate course where they would learn to apply the guidelines for depression. We aimed for a minimum of 40 and a maximum of 50 participants working in two manageable workgroups. To participate, IPs had to be available on eight different dates for the purposes of randomisation and allocation, which created difficulties. We used various strategies for the recruitment of IPs to participate in our study. We started at the Institute by giving oral presentations at five of the 23 branch offices (Utrecht, The Hague, Rotterdam, Venlo and Heerlen), by mailing all IPs (circa 900) at their Institute addresses and by advertising on the internal website of the Institute. Furthermore, we asked the general staff, the staff IPs and the local managers of the Institute to support participation in the study. During this period of recruitment the following reasons for non-participation were noted: no need for this course, no permission from the manager, too busy with



practising and no time left for education, not interested, limited allowance for attending courses, preference for other courses, too much travelling time and unknown. By January 2009, we had 21 IPs interested in participation. However, when the planned dates for the course were announced, only seven IPs remained, i.e. not enough to participate in the study. We then cooperated with the Netherlands School of Public and Occupational Health (NSPOH) and included our study in a post-graduate course for IPs located at the NSPOH. We then sent a newsletter promoting participation to all registered and trainee IPs (797 altogether) on the NSPOH mailing list. Furthermore, the post-graduate course was accredited by the NSPOH as an incentive. In March 2009, just before the planned start of the course, another 36 IPs signed, finally reaching a total of 43. Seven out of approximately 900 and 36 out of 797 (the mailings lists of the Institute and the NSPOH largely coincide with each other) resulted in a reach of approximately 5%.

**Table 8.2:** Baseline characteristics of participating IPs (n=42).

<b>IP characteristics</b>	
Age (mean $\pm$ sd years)	50.5 (5.8)
Male (%)	52.4
<b>Work related characteristics</b>	
Registered as IP (%)	85.7
Years working as IP (mean $\pm$ sd)	15.4 (7.8)
Working hours/ week (mean $\pm$ sd)	31.7 (9.3)
Number of clients with depression assessed per month (mean $\pm$ sd)	7.0 (5.0)
<b>GD related characteristics</b>	
In possession of GD (%)	92.9
IPs read GD (%)	
Most of it	28.6
Whole	69.0
IPs used GD in practice (%)	85.7
Number of GD consultations (%)	
0	4.8
1 – 5	57.1
6 – 10	26.2
< 10	11.9

IP=Insurance physician; GD=guideline Depression.

These 43 IPs were randomized and allocated to the intervention or to the control group of the experiment. The participants completed the baseline questionnaire providing demographic information, work-related characteristics and guideline-for-depression characteristics. The baseline characteristics of the participating IPs are shown in Table 8.2. Data from the Institute showed that the group of participating IPs was representative of the total population of IPs working at the Institute in terms of age, gender, registration as an IP and working hours per week. The mean age of the total population of IPs (N = approximately 900) was 49 years, 58.3% were male, 85% were registered and they worked for on average 32 hours per week [18].

**Table 8.3:** The implementation strategy: different components of the training day and the evidence-based tools.

#### **Components of the training day, educational strategy, time needed**

Introduction - plenary - 0.15 hr

Presentation by a psychiatrist - interactive presentation with the complete group - 1 hr

Trainer instructing on tools and on focus on videotape of client played by an actor - 6 subgroups, interactive, coaching - 0.30 hr

IPs filling in checklist 1 and 2 while watching the videotape of client played by an actor - self-activity - 1 hr

Trainer translating GD into physician's practice - 6 subgroups, interactive - 0.30 hr.

IPs in subgroups assessing disability of the taped client played by an actor using the tools and presenting their findings to the complete group - self-activity, feedback by trainer - 1 hr

IPs in subgroups write down their considerations of the assessment report by using the essential elements of reasoning - self-activity, feedback, coaching - 1 hr

#### **Evidence-based tools**

Plastic summary card: diagnostics, co-morbidity, therapy, treatment, ICF-model

Checklist 1: main items of the GD

Checklist 2: Disabilities connected to the severity of the depression

Hamilton Rating Scale for Depression [14]

GD=guideline Depression.

#### **Dose delivered and dose received**

The training day consisted of four main components. Various educational strategies, such as interactive presentations, self-activation, feedback, subgroup presentations and individual coaching, were used in order to engage the participating IPs in applying the guidelines. They were thoroughly and individually coached by a trainer on how to use the evidence-based tools for the assessment of a client with depression. They practised their skills in assessing work disability with help of a realistic videotape of a client with

depression, played by an actor, and received immediate feedback on their performances. The training day was planned according to a strict time schedule. The research assistant was responsible for the schedule and reported via a checklist that the training was delivered exactly as was planned in the implementation program (see table 8.3).

Between randomisation and baseline measurement one of the 43 IPs withdrew because of lack of availability; 42 (98%) completed the course including the training with the tools. The questionnaires for the process evaluation were filled out during the course, so the rate of response was also 98%. In general, the participants showed engagement with the training, participated actively, used the tools and responded well to feedback, as was mentioned in the evaluation report made by the research assistant immediately after the training day.

### Participant satisfaction and expectations

Overall, the IPs were satisfied with the training day, rating it with an average score of 7.7 (sd = 0.6; median = 8.0) on a 1-10 scale. The different components of the training program were rated highly, between 4.2 and 4.7 on a 1-5 scale (see Table 8.4).

**Table 8.4:** Satisfaction with the different components of the training (n=42).

Components of the training day	Mean score† (sd), median
Presentation by a psychiatrist	4.7 (0.5), 5.0
Trainer translates GD into physician's practice	4.3 (0.6), 4.5
Assessing disability of a client presented on video, interactive and in subgroups and training to use the evidence-based tools	4.2 (0.5), 4.3
Exercising in writing the assessment report by using the essential elements of reasoning	4.2 (0.9), 4.5

†1-5 scale, 5 indicating maximum; GD: guideline Depression; Cronbach's Alpha = 0.788.

Directly after the training day, 81.0% of the IPs expected improvement in their disability assessments. Furthermore, 85.8% expected improvement in their assessment reports with the use of the tools handed out during the course. Of all tools, IPs had the highest intention of using the plastic summary card (4.8 on a 1-5 scale), followed by the checklists 1 and 2 (rated 4.3 and 4.2), whereas the HRSD [13] was rated lowest (3.3). Furthermore, the evidence-based tools were expected by IPs to be most useful for making decisions about the diagnosis and were rated with 4.6 on a 1-5 scale, in an order followed by estimating the severity of depression (4.5), assessing the client's history, risk factors and progress (4.1), co-morbidity (4.1), treatment and convalescence (4.1) and determining work limitations (4.1).

After three months 95.7% of the IPs (n=23) still considered the training as having been useful. Despite this, 17.4% wanted more training in using and applying the



guidelines for depression. After three months they were still most satisfied about the plastic summary card (4.7 on a 1-5 scale) and less satisfied about the HRSD [13] (3.0). Checklist 1 was rated as 4.1 and checklist 2 as 3.9; three months after having received the training satisfaction with the guidelines was 4.3 on a 1-5 scale.

**Table 8.5:** Perceived barriers to the use of evidence-based tools after the training day (n=42).

Aspect	Factor	Mean score† (sd), median
Professional (IPs)	Expertise	2.2 (1.1), 2.0
	Work style	2.6 (1.2), 2.0
	Attitude	1.8 (1.0), 2.0
	Doubts innovation	2.2 (1.0), 2.0
Innovation (evidence-based tools)	Perceived advantage	1.8 (0.7), 1.8
	Time-investment	2.5 (1.0), 2.3
	Compatibility	2.1 (0.7), 2.0
	Scientific basis	2.1 (0.8), 2.0
	Validity	1.7 (0.7), 2.0
	Feasibility	1.6 (0.7), 2.0
	Flexibility	2.1 (0.6), 2.0
	Transparency	1.7 (0.6), 2.0
	Didactive benefit	1.5 (0.6), 1.5
	Attractiveness	1.5 (0.7), 1.0
	Applicability	1.6 (0.8), 1.0
	Complexity	2.1 (1.1), 2.0

†Scale ranged from no barrier perceived (1) to barrier perceived (5); IP=Insurance physician; Cronbach's Alpha Professional = 0.704; Cronbach's Alpha Innovation = 0.743.

### Perceived barriers

Barriers to the use of the guidelines with the tools were explored at the level of the professional (IPs) and at the level of the innovation (evidence-based tools). Table 8.5 gives an overview of the most important barriers perceived by the IPs directly after the training day. The higher the score, the more it was perceived as a barrier. At the level of the professional, the work style (difficulties in changing routines) of the IP was the perceived barrier with the highest rating (2.6 on a 1-5 scale). At the level of the tools the factor of acquired time for using the tools to apply the guidelines was the perceived barrier with highest rating (2.5 on a 1-5 scale). Other aspects of the tools, such as

educational benefits, applicability, attractiveness and transparency were not perceived as barriers. These aspects showed that the tools themselves appeared to suit to the needs of the IPs.

After three months, there were two items which were rated significantly differently by IPs compared with the rating directly after the training: their attitudes to both the innovation and the validity of the tools changed in a positive way. Ratings of all other barriers did not change (see Table 8.6). Perception of work style and time investment remained a barrier and even increased.

**Table 8.6:** Perceived barriers to the use of the tools after the training day and at three-month follow-up (n=23).

Aspect	Factor	Mean score† (sd), median	
		Directly after training	At three-month follow-up
Professional (IPs)	Expertise	2.4 (1.2), 3.0	2.0 (1.0), 2.0
	Work style	2.7 (1.2), 3.0	2.8 (1.3), 3.0
	Attitude	1.7 (0.7), 2.0	1.7 (1.1), 1.0
	Doubts innovation	2.1 (1.2), 2.0	1.7 (0.9), 1.0*
Innovation (evidence-based tools)	Perceived advantage	1.7 (0.7), 1.5	1.7 (0.6), 2.0
	Time-investment	2.7 (1.1), 3.0	2.8 (1.2), 3.0
	Compatibility	2.0 (0.7), 2.0	2.1 (1.0), 2.0
	Scientific basis	2.2 (0.7), 2.0	2.0 (0.8), 2.0
	Validity	1.8 (0.7), 2.0	1.6 (0.7), 2.0**
	Feasibility	1.6 (0.7), 1.0	1.8 (1.0), 2.0
	Flexibility	2.0 (0.6), 2.0	2.0 (0.6), 2.0
	Transparency	1.7 (0.6), 2.0	1.7 (0.8), 2.0
	Didactic benefit	1.4 (0.5), 1.0	1.5 (0.6), 1.0
	Attractiveness	1.6 (0.8), 1.0	1.6 (0.8), 1.0
	Applicability	1.9 (0.9), 2.0	2.0 (1.1), 2.0
	Complexity	2.0 (1.0), 2.0	1.7 (1.0), 1.0

†Scale ranged from no barrier perceived (1) to barrier perceived (5); IP: insurance physician; Cronbach's Alpha Professional After training = 0.689, At three-month follow-up = 0.620; Cronbach's Alpha Innovation After training = 0.773, At three-month follow-up = 0.850; doubts innovation n = 21, applicability n = 22, complexity n = 21; \* p = 0.021; \*\* p = 0.043.



## Discussion

### Main findings

The aim of this study was to evaluate the newly developed implementation strategy for the guidelines for depression. The results of this process evaluation show that despite intensive recruitment efforts only a small group of 43 IPs were reached. However, 42 of this group received all the components of the implementation strategy as planned. Changing their work routine and finding the time to apply the guidelines were found to be barriers to the implementation strategy, but most IPs expected the course to improve their assessments of clients with depression. Implementation and fidelity to the implementation strategy in practice could not be measured because the design of this study was an experiment in a controlled setting.

### Comparison with other studies

The low reach of this study was comparable to other public health studies involving the recruitment of physicians [19-21]. However, the fact that the participants had to be available on eight separate dates could be a serious barrier to participation. The IPs who did participate might be self-selective and motivated for learning to apply the guidelines for depression. In the experiment in a controlled setting, the IPs who received the implementation strategy performed significantly better on implementing the guidelines for depression than did the control group [8]. This process evaluation assesses various aspects of the implementation strategy and shows that IPs were satisfied with the training and tools and that the majority of them had positive expectations about the use of the guidelines in practice. Satisfaction should be regarded as an important process measure because it has been shown to be closely associated to the acceptance of and participation in the provided service, in this case adherence to the guidelines for depression [22].

In a systematic review evaluating the effect of clinical guidelines the conclusion was made that, in the context of rigorous evaluations, explicit guidelines do improve clinical practice [23]. However, another review by the same authors states that the successful introduction of clinical guidelines is dependent on many factors including the clinical context as well as the methods of developing, disseminating and implementing the guidelines [24]. Poor implementation often hampers the impact of practice guidelines on quality of care [25]. Guideline adherence depends on internal barriers (i.e. lack of awareness, lack of familiarity, lack of agreement, lack of self-efficacy, lack of outcome expectancy and the inability to overcome the inertia of previous practice) and on external barriers (i.e. patient, environmental and guideline factors) [26].

From earlier studies it is known that IPs in general show positive attitudes towards evidence-based medicine guidelines [27-29] although their use of guidelines was in practice rather poor. Our developed implementation strategy aimed to diminish this gap



between attitude and performance by supporting the IPs with training and tools, making it easier for them to conform to the guidelines. The only internal barrier mentioned in guideline adherence by the IPs in our study was their perceived problems with changing their routines which use of the guidelines would imply. However, recognizing changing routines as a problem is closer to adherence to guidelines than is rejecting them, e.g. because of lack of agreement or because they consider them to be 'cook-book' medicine.

One perceived external barrier to the use of the guidelines was the time factor; guideline recommendations that require extra time investment are less adhered to than those that can save time [30]. Lack of time is also known to be one of the main barriers to practising evidence-based guidelines for occupational physicians and general practitioners [31, 32]. The IPs' concerns about the extra time needed to use the guidelines should have been addressed in the training. In fact, the tools were developed with the intention of facilitating the implementation of the guidelines by working more efficiently than usual. Unfortunately, our implementation strategy did not completely eliminate the idea that the use of guidelines involves extra time. Indications for barriers to the translation and transfer of knowledge from the guidelines to the IPs were not found in this study; this seems to show that this implementation strategy covers the knowledge infrastructure that is needed for evidence-based decision making in insurance medicine [33].

Training sessions alone are not likely to be effective in changing complex behaviours such as a physician's guideline adherence. Strategies to increase attendance at such sessions, the use of mixed interactive and didactic formats and placing the focus on outcomes perceived as serious - as was done in our programme - may all increase the effectiveness of training sessions [34].

### Strengths and limitations of the study

No other study has evaluated the implementation of insurance medicine guidelines in an experimental setting. In the controlled setting of the post-graduate course we managed to deliver the strategy as planned. This process evaluation contains extensive data on the implementation strategy as delivered and perceived. By using IP-trainers who had broad experience of the work in practice, the translation of evidence-based guidelines to the IPs' practice worked out well. Feedback given by the trainers on applying the guidelines using realistic case histories presented on video connected evidence-based medicine with practice. Dose received was measured directly at the training session. Another strength was that, for the collection of data to measure the perceived barriers, we used the WOK questionnaire [17] and achieved a 98% response rate. Furthermore, this process evaluation was considered to be an essential part of the design of the experiment and was developed early in the planning [35].

A limitation of this study was that we reached only 5% of the total group of IPs, which might indicate a selection bias. According to the Linnan and Steckler framework

[16], we have no results on the element implementation, because the implementation strategy in this study has only been executed in a specific controlled setting and not yet in actual practice. That could be the next step in further research on this subject. Although we have subjective impressions and reports on the element fidelity of this study, fidelity was not objectively measured. Immediately after the training the IP-trainers reported that the participants showed engagement with the implementation strategy but proof of that is lacking. Furthermore, we should be aware that our data were obtained in a controlled setting with motivated IPs, which means that not all results and conclusions can be generalized to the complete group of IPs working in practice. There might be barriers to the implementation strategy at organizational or client level which remain unknown.

### Implications for practice

The developed implementation strategy could be distributed and evaluated on a wider scale than the limited reach of this experiment. For instance the Institute where most of the IPs in the Netherlands work could support this implementation strategy by including the one-day training in their internal education program.

### Conclusions

This process evaluation shows that the developed implementation strategy can be successful in a controlled setting. Reach was poor, but dose delivered, dose received and satisfaction with the implementation strategy itself were good, even after three months. Most of the IPs expected to improve their disability assessments of clients with depression. In this implementation strategy two perceived barriers were identified: changes to work routines and the time required for using the guidelines.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

The authors declare that they participated in the study and made the following contributions to the study, and that they have seen and approved the final version. FZ, AJMS, JRA, and AJvdB contributed to the conception and design of this study. KG and FZ contributed to the analysis. FZ and KG wrote the manuscript. AJMS, JRA, and AJvdB revised and commented on the manuscript. AJMS, JRA and AJvdB will act as guarantors of this study.



## Acknowledgements and funding

Funding sources: FZ, JRA and AJMS are (partially) funded by UWW. The study sponsor had no decisive role in the study design, in the collection, analysis, or interpretation of the data, or in the decision to submit the paper for publication.

## References

1. Bijl RV, Ravelli A, Van Zessen G: Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Soc Psychiatry Psychiatr Epidemiol* 1998, **33**:587-95.
2. Murray CJ, Lopez AD: Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet* 1997, **349**:1498-1504.
3. Lagerveld SE, Bultmann U, Franche RL, Van Dijk FJ, Vlasveld MC, Van der Feltz-Cornelis CM, et al.: Factors associated with work participation and work functioning in depressed workers: a systematic review. *J Occup Rehabil* 2010, **20**: 275-292.
4. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al.: Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 1999, **282**:1458-65.
5. Spanjer J, Krol B, Brouwer S, Groothoff JW: Sources of variation in work disability assessment. *Work* 2010, **37**:405-11.
6. Health Council: Assessment, treatment and counselling. *Medical responses to sickness absenteeism and occupational disability*. Den Haag: Gezondheidsraad, 2005.
7. Health Council: *Insurance Medicine Guidelines for Depression*. Den Haag: Gezondheidsraad, 2006.
8. Zwerver F, Schellart AJ, Knol D, Anema JR, Van der Beek AJ: An implementation strategy for insurance physicians to improve guideline adherence: an experiment in a controlled setting. *BMC Implementation Science*, 2011, **6**:131.
9. Grol R: Successes and failures in the implementation of evidence-based guidelines for clinical practice. *Med Care* 2001, **39**:II46-II54.
10. Grol R, Grimshaw J: From best evidence to best practice: effective implementation of change in patients' care. *Lancet* 2003, **362**:1225-1230.
11. Hulscher ME, Laurant MG, Grol RP: Process evaluation on quality improvement interventions. *Qual Saf Health Care* 2003, **12**:40-46.
12. Oakley A, Strange V, Bonell C, Allen E, Stephenson J: Process evaluation in randomised controlled trials of complex interventions. *BMJ* 2006, **332**:413-416.
13. Hamilton M: A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960, **23**:56-62.
14. World Health Organization. *ICF: International Classification of Functioning, Disability and Health*. Geneva: WHO, 2001.
15. Slebus FG, Kuijjer PP, Willems JH, Frings-Dresen MH, Sluiter JK: Work ability in sick-listed patients with major depressive disorder. *Occup Med* 2008, **58**:475-9.
16. Linnan L, Steckler A: Process evaluation for public health interventions and research: an overview. In: *Process evaluations for Public Health Interventions and Research*, San Francisco, CA. *Jossey-Bass* 2002, 1-23.
17. Peters M, Harmsen M, Laurant M, Wensing M: *Room for improvement? Barriers to and facilitators for improvement of patient care*. Nijmegen: Centre for Quality of Care Research (WOK), Radboud University Nijmegen Medical Centre, 2002.

18. Steenbeek R, Schellart AJ, Mulders H, Anema JR, Kroneman H, Besseling J: The development of instruments to measure the work disability assessment behaviour of insurance physicians. *BMC Public Health* 2011, **11**:1.
19. Rebergen DS, Bruinvels DJ, Bezemer PD, Van der Beek AJ, Van Mechelen W: Guideline-based care of common mental disorders by occupational physicians (CO-OP study): a randomized controlled trial. *J Occup Environ Med* 2009, **51**:305-12.
20. Sinnema H, Franx G, Volker D, Majo C, Terluin B, Wensing M, et al.: Randomised controlled trial of tailored interventions to improve the management of anxiety and depressive disorders in primary care. *Implement Sci* 2011, **6**:75.
21. Verweij LM, Proper KI, Hulshof CT, Van Mechelen W: Process evaluation of an occupational health guideline aimed at preventing weight gain among employees. *J Occup Environ Med* 2011, **53**:722-9.
22. Guidotti T, Cowell J, Jamieson C: *Program evaluation*. In: American Medical Association. *Occupational health services, a practical approach*. Chicago: AMA, 1989.
23. Grimshaw JM, Russell IT: Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993, **342**:1317-22.
24. Russell IT, Grimshaw JM: The effectiveness of referral guidelines: a review of the methods and findings of published evaluations. In: Roland M, Coulter A, eds. *Hospital referrals*. Oxford University press, 1992: 179-211.
25. Grilli R, Lomas J: Evaluating the message: the relationship between compliance rate and the subject of a practice guideline. *Med Care* 1994, **32**:202-13.
26. Cabana MD, Ebel BE, Cooper-Patrick L, Powe NR, Rubin HR, Rand CS: Barriers pediatricians face when using asthma practice guidelines. *Arch Pediatr Adolesc Med* 2000, **154**:685-93.
27. Heselmans A, Donceel P, Aertgeerts B, Van de Velde S, Ramaekers D: The attitude of Belgian social insurance physicians towards evidence-based practice and clinical practice guidelines. *BMC Fam Pract* 2009, **10**:64.
28. Heselmans A, Donceel P, Aertgeerts B, Van de Velde S, Ramaekers D: The attitude of Flemish occupational health physicians towards evidence-based occupational health and clinical practice guidelines. *Int Arch Occup Environ Health* 2010, **83**:201-8.
29. Kok R, Hoving JL, Verbeek JH, Schaafsma FG, Smits PB, Van Dijk FJ: Evaluation of a workshop on evidence-based medicine for social insurance physicians. *Occup Med* 2008, **58**:83-7.
30. Van den Berg MJ, De Bakker DH, Spreeuwenberg P, Westert GP, Braspenning JC, Van der Zee J, et al.: Labour intensity of guidelines may have a greater effect on adherence than GPs' workload. *BMC Fam Pract* 2009, **10**:74.
31. Hugenholtz NI, Nieuwenhuijsen K, Sluiter JK, Van Dijk FJ: Do knowledge infrastructure facilities support evidence-based practice in occupational health? An exploratory study across countries among occupational physicians enrolled on Evidence-Based Medicine courses. *BMC Health Serv Res* 2009, **9**:18.
32. Karbach U, Schubert I, Hagemeister J, Ernstmann N, Pfaff H, Hopp HW: Physicians' knowledge of and compliance with guidelines: an exploratory study in cardiovascular diseases. *Dtsch Arztebl Int* 2011, **108**:61-9.
33. Van Dijk FJ, Verbeek JH, Hoving JL, Hulshof CT: A knowledge infrastructure for occupational safety and health. *J Occup Environ Med* 2010, **52**:1262-8.
34. Forsetlund L, Bjorndal A, Rashidian A, Jamtvedt G, O'Brien MA, Wolf F, et al.: Continuing education meetings and workshops: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2009, CD003030.
35. Bartholomew LK, Cushman WC, Cutler JA, Davis BR, Dawson G, Einhorn PT, et al.: Getting clinical trial results into practice: design, implementation, and process evaluation of the ALLHAT Dissemination Project. *Clin Trials* 2009, **6**:329-43.

# **Chapter**

---

**9**

## **General discussion and conclusions**

The aim of this thesis was to develop and to evaluate an implementation strategy for (one of the) insurance medicine guidelines that suits to insurance medicine practice. At the start of this research project (the end of 2006) evidence-based medicine was only recently introduced in insurance medicine by guidelines. Between 2007 and 2009 20 guidelines, specific for insurance medicine, were developed and implemented at high pace at the Dutch Institute for Employee Benefits Schemes (Institute). At that time, there was no experience with implementing guidelines. Research concerning the implementation of guidelines in the field of insurance medicine was needed. Therefore, an implementation strategy for the guidelines for depression was developed and evaluated in a controlled experiment. The questions put in the General Introduction will be answered now. Furthermore, methodological considerations will be discussed, and our developed implementation strategy will be compared to the implementation of medical guidelines in general. Practical implications for IPs, Institute stakeholders, and education programmes in insurance medicine will be discussed. Finally, recommendations for further research will be given.

## Answers to the questions and main findings

### **Which strategy can be developed to implement the guidelines for depression, in order to promote use by IPs?**

A multifaceted implementation strategy, for the implementation of the guidelines for depression could be developed. An essential element of this multifaceted strategy was a tailor-made training, in which the IPs, facilitated by tools, learned to apply the guidelines for depression. The intervention mapping (IM) method facilitated integrating opinions and needs of IPs and experts with theories, to support the development of the strategy [1].

### **What are the effects of such a developed implementation strategy on the guideline adherence of the IPs and on their knowledge of the guidelines?**

The IPs who received the implementation strategy adhered more to the guidelines for depression than the IPs from the control group. The IPs who received the implementation strategy showed more knowledge of the guidelines for depression, than the IPs from the control group.

### **What are the effects of such a developed implementation strategy on the behavioural determinants of the IPs regarding the use of the guidelines?**

The investigated determinants of the IPs' behaviour concerning the guidelines changed positively after having received the implementation strategy. Their attitude towards



the guidelines improved. The IPs' self-efficacy and intention for using the guidelines increased. Furthermore, they reported improved knowledge and skills with regard to the guidelines.

### **What are the effects of such a developed implementation strategy on the inter-IP agreement in the work disability assessments of the IPs?**

The implementation strategy caused an increase in the inter-IP agreement of work disability assessments of clients with depression. This implies improvement of uniformity in the work disability assessments due to the improved guideline adherence.

### **What are the effects of such a developed implementation strategy on the number and severity of work limitations when applying the guidelines?**

The IPs who had received the implementation strategy tended to assess more often and more severe work limitations for the same client cases than the IPs from the control group.

### **What are the effects of such a developed implementation strategy on the satisfaction of the IPs?**

Overall, IPs were highly satisfied with the implementation strategy. In particular, they appreciated the training and one of the tools, the summary desk mat. They expected to improve their assessments of clients with depression by applying the guidelines. Time needed for applying the guidelines sometimes was mentioned as a barrier to the implementation, and some of the IPs had concerns about changing their work routines for applying the guidelines.

## **Methodological considerations**

In the previous chapters methodological strengths and limitations regarding each respective chapter were described. Some additional methodological and practical considerations will be discussed here.

## **Development of the implementation strategy**

When planning the project, the Intervention Mapping method (IM) [1] was chosen because this method has shown to be useful for the development of theory- and practice-based interventions aimed at return to work of sick-listed employees [2, 3]. In this research project it was the first time that IM has been used for the development of an implementation strategy for guidelines in the field of insurance medicine. The IM method supplies a framework, consisting of six steps, in which the opinions of IPs, staff-

IP, managers, IP trainers, psychiatrists and researchers could be integrated with theories, providing a qualitative and sound development of the implementation strategy. A strong point was that the findings of the heterogeneous group of experts, were adapted on base of a needs assessment to the needs of IPs in practice. IM provides not only the development of the implementation strategy, but also takes care of the implementation and evaluation of the strategy, which suits the context of implementing guidelines. In our study the evaluation of implementing guidelines was limited mainly to the level of the physician, who moreover acted in an experimental setting. Even then, it appeared that IM was a complex and time-consuming method. For saving time we probably could have left out some of the six IM steps. However, at the start of our project, we could not estimate whether other methods, such as a review, alone or in combination with focus groups, would have produced the same results more efficiently than the IM method. IM as a method seemed to be useful for the development of an implementation strategy for guidelines, because the development of an approach for implementing guidelines into practice has been described as a complex process for which evidence and feasibility often lacks [4].

Finding the right approach was of particular importance, because the field of insurance medicine at the Institute, where the guidelines were implemented, is characterized by competing interests. IPs have an interest in their professional autonomy and quality of work. Institute stakeholders have an interest in the management of work disability assessments regarded as a process of production qualified by quantity and timeliness. Furthermore, national politics with its proceeding legislation have the scope laid on work disability benefits as a societal problem.

These contextual aspects, which are specific for the field of insurance medicine at the Institute, could probably not have been covered by a literature review alone or a Delphi method. Additionally, by using the IM method, which includes bottom-up elements, our approach differed fundamentally from the usual implementation of guidelines at the Institute, which rather was a top-down approach. Although, from research point of view, the effectiveness of the complete IM method has not sufficiently been proven yet.

### **The design of the study**

In line with the IM method we planned to evaluate the developed implementation strategy in practice. The original idea was to realise this aim by an effectiveness study with the design of a randomised controlled trial (RCT). Unfortunately, the effectiveness study failed for practical reasons. The needed number of participants (200) could not be reached in the limited period for recruitment activities. We managed to recruit only 43 participants for our study. Therefore, the original plan of the effectiveness study was left for an efficacy study. The efficacy design requires fewer participants compared to an effectiveness study. The efficacy design provided us with the opportunity of having the participants assessing the same client cases, which is hard to realize in real practice.



The efficacy design has some other advantages compared to an effectiveness study. By carrying out the efficacy study in a controlled setting and a short timeline we could secure optimal compliance of the IPs, resulting in a high response rate and a minimal loss to follow-up. Opposite to real practice, selection bias of the assessed cases by the IPs was not possible in the controlled experiment. Another feature of the efficacy design was, that the study was independent of organizational factors, such as lack of support by local managers at the Institute or reorganizations leading to loss of follow-up. Furthermore, the short timeline of the study enabled us to minimize the external influences on the participants and avoid mutual contamination between the groups. On the other hand, due to this short timeline, the number of measurements was limited. Hence, long-term effects of the implementation strategy could not be investigated.

However, one might argue that for evaluating the developed implementation strategy an effectiveness study is preferable to an efficacy study, because the effectiveness study should be carried out in practice and the efficacy study took place in a controlled setting. We evaluated the implementation strategy in specific laboratorial setting, and we do not know whether the implementation strategy works in practice as well as in the controlled setting.

One important difference of our efficacy study compared to an effectiveness study was that in our study the guidelines for depression were applied to cases of depression played by actors and presented on a video screen, instead of real clients. Therefore, we could not evaluate the developed implementation strategy at the level of client outcomes, e.g. their satisfaction with the guidelines for depression. In the controlled experiment the IPs made work disability assessment reports after watching the video cases of simulated clients with depression. The IPs could not ask questions to the simulated clients, like they are used to do in practice. Video cases have been used in medical education many times. Video cases can promote enjoyable learning and appeared to be valuable for group discussions [5-7]. Although the IPs perceived the video cases as realistic as their own practice, the use of video cases instead of real clients remains a methodological limitation in this study because video cases can only be compared to real clients under certain conditions. Unfortunately, in insurance medicine an optimal client modality does not exist, such as a golden standard of the work disability assessment of that certain client, which limits generalization of our results to real practice. This methodological limitation is not exclusive for insurance medicine, it even occurred in the comparison of the assessment of cardiac physical examination skills by internists between simulation technology and real patients [8]. If the study had been an effectiveness study and had taken place in real practice, then the results might have been more distinct compared to this experimental setting, because in real practice the IPs could have had influence on the interview part of the assessment, as well as on the written part of the assessment. In real practice they could have applied their learned skills by asking the right questions to the client.

However, an effectiveness study carried out at the Institute could have had its own drawbacks except from its lack of practical feasibility. The IPs might not see enough clients with depression, given a certain period. IPs might not have enough time, due to production requirements, for compliance to the research protocol, and there could have been different local influences of management or staff on the IPs, dependent on the front office where they work.

No matter which design, a selection bias would have occurred because we wanted voluntary participants in the study. The controlled experiment met the CONSORT statement for trials: randomisation, allocation, procedures of blinding, participant flow, sample size, intention-to-treat and per-protocol analysis, and report of all outcomes were performed as required by the CONSORT statement [9]. As the number of measurements was limited, long-term effects of the implementation strategy could not be evaluated.

### **Primary outcome: guideline adherence and measurement instrument**

Guideline adherence of insurance physicians can be measured directly by observing their performance in the consultation room, or indirectly by measuring a derivative of their performance. We used the IPs' work disability reports for measuring their guideline adherence, assuming that guideline adherence can actually be measured by evaluating a work disability report of an IP. Direct methods would be difficult to realize for reasons of medical secrecy and IPs' limited compliance to research projects at the Institute. Clients could be replaced by actors in the consultation room, playing cases of depression. However, that option is inefficient for planning in practice at the Institute because of huge logistic demands. Besides, using actors could have interfering influences (confounding effects) on the outcome guideline adherence, because none of the work disability assessments would then proceed identical. Furthermore, the costs of using actors for the measurements would exceed the available budget.

For measuring guideline adherence we developed performance indicators (PI), since another measurement instrument for insurance medicine guidelines did not exist. These PIs were developed by IPs, expert IPs and the researchers. Firstly, the PIs had to show good content validity, secondly the PIs were tested for reliability, and finally the PIs enabled a scoring method for our data. These PIs are more than a list of items. These PIs are not just process indicators. These PIs actually measure adherence of the IPs on applying guidelines. The format of the PIs exists of six decision trees, providing the opportunity for making the 'mind lines' or IPs' logic paths of reasoning visible to the judge. Additionally, the PIs suit to educational aims, such as providing feedback to an IP on his/her performance on applying the guidelines for depression. Furthermore, PIs are an essential step in the process of guideline implementation by monitoring performance of IPs continuously and providing feedback about this.

It should be remarked that only guideline adherence in the written part of the disability assessment by an IP could be measured by the PIs, not in the interview part



of the assessment. On the one hand, however, it seems plausible that a work disability report with the maximum of adequate scores at the PIs really reflects a work disability assessment that has been carried out in concordance with the guidelines, because the PIs reflect all main elements of the guidelines, including the paths of decision making by an IP. On the other hand, when the PIs will be used in practice, it is possible that an IP has carried out a work disability assessment in concordance with the guidelines but yet the report shows inadequate scores at the PIs. In this case, the IP has correctly followed the guidelines in the interview part of the assessment, but subsequently failed in reporting these findings in concordance with the guidelines. This might occur especially when an IP is working under time pressure due to production requirements. As is shown in Chapter 8, IPs perceive that writing a disability report in full concordance with the guidelines takes more time than usual. The other way round is also possible, an IP's assessment report shows only adequate scores, while the assessment might have been insufficient. This happens, for example, when an IP tends to follow his or her own thoughts, such as a tunnel vision. In that case all information gathered during an assessment is used to confirm their first impression of the client and other relevant information that could have led to a different diagnosis or work limitations is judged as less relevant. That could lead to adequate PI scores, violating reality. This phenomenon refers to one of the potential barriers for using guidelines, which is called 'cook-book medicine'. Guidelines should be used carefully and with respect to professionals' autonomy [10].

### **The implementation strategy**

Essential is that the implementation strategy in this research project was developed on base of the IPs' needs. The IPs' perceived difficulties in applying guidelines in practice after the usual implementation of guidelines at the Institute. In semi-structured interviews with trainer IP experts, the training was designed and later performed by two trainer IPs. This training including the facilitating tools (i.e. the developed implementation strategy) was positively evaluated by the participating IPs. Regarding the results of the efficacy study, we know that the intervention worked in the controlled setting, but we do not exactly know which parts of the training or the tools caused the effects. The training lasted only one day, which is rather short. Preferably, it would have been better to give the IPs more opportunity to bring their learned skills in practice, and then evaluate their performances in practice at another planned day.

On closer inspection, our implementation strategy contains elements based on three of the five approaches mentioned by Grol [4]. Firstly, the approach of evidence-based medicine, assuming that provision of best evidence and convincing information leads to optimal decision making and optimal use of guidelines. Secondly, the educational approach, which is characterized by an internal striving for professional competence of the physicians. In small groups the physicians learn from their own experiences, interactively, and feel that they own the required changes. Thirdly, the approach of



assessment and accountability, providing feedback on performance relative to peers, motivating change in practice routines. The remaining two approaches were not covered by our implementation strategy. Our implementation strategy lacked patient involvement or shared decision making, and was not aimed at total quality management, restructuring processes or quality systems. This implementation strategy was limited to the level of the professionals.

Although we initially used a psychological behavioural model as a starting point for this study, some of its aspects regarding influencing IPs' behaviour became part of the developed implementation strategy. By translating the guidelines for use in practice, the use of tools, and giving the IPs feedback on their performance, we aimed to influence the IPs' behaviour towards the guidelines. Furthermore, we paid attention to social interaction by using respected opinion leaders and IP trainers with a broad experience in insurance medicine practice for enforcing and stimulating our implementation strategy. If the implementation strategy will be carried out in real practice, we can utilize the power of social interaction even more, by involving stakeholders, staff IPs, and regular IP group meetings in the strategy.

### Exploration of the behaviour of IPs regarding guidelines

The psychological behavioural model ASE model as a derivative from the Theory of Planned Behaviour (TPB) gave support to describe IPs' behaviour towards guidelines [11, 12]. This model has been successfully used for describing behaviour and changes in behaviour in the field of occupational health and also in insurance medicine [3, 13, 14]. This model as integrated in IM was used for the development of the implementation strategy. Although validated questionnaires for measuring ASE determinants could be identified in literature concerning clinical care, primary care and occupational health care, we had to transpose the questions to the field of insurance medicine.

We used the ASE model as a starting point for describing the IPs' behaviour towards the implementation of guidelines in general and more specific to the guidelines for depression. In a review concerning the use of TPB models for understanding physician's behaviour, it was concluded that understanding physicians' determinants of behaviour, such as attitudes, social norms, and self-efficacy, and providing the necessary support are the keys to develop a strategy that is most likely to impact behaviour [15]. Therefore, we developed an implementation strategy, a multifaceted educational programme, which might realize this aim. We hypothesised that the implementation strategy might have a positive impact on all ASE determinants, such as attitude, self-efficacy, knowledge and skills, and intention as well as on their relationships with intention to use the guidelines or with use of the guidelines [16].

In an article concerning testing the effectiveness of TPB interventions targeting professional's guideline adherence, the authors mentioned that these intervention studies (such as our study is) can be criticized on three methodological grounds [17].

They indicated 1) the aspect of over-reliance on self-reported outcomes; 2) the failure to control for influences on behaviour other than the intervention; and 3) the short follow-up periods making it difficult to establish the longer-term impact of the interventions. In our study, we addressed the first two concerns. The results of our investigations of the IPs' behaviour on base of the ASE model were indeed self-reported. However, we have linked these self-reported results with objectively assessed written disability assessment reports using PIs (Chapter 6). Furthermore, we could reduce influences on the IPs' guideline adherence other than the implementation strategy, because of the chosen design, which was an experiment in a controlled setting. The third concern, the limited period of follow-up, was a serious drawback indeed. We performed two measurements, at baseline and at three months of follow-up after the intervention. Preferably we should have measured changes in behaviour directly after the intervention, and at three months as well.

## Considerations concerning the implementation of guidelines

### Effective implementations

Grol, Wensing and Eccles presented a box containing elements of effective implementations [18]. In Table 9.1 the features of our developed implementation strategy are pointed out with a +, or - sign.

Looking at the features in Table 9.1, it appears that our developed implementation strategy meets the majority of the required elements for an effective implementation. Regarding the fourth element, this was given a negative score because in our group of participants we did not distinguish subgroups. We could have made subgroups e.g. according to stages of change, or years of experience as an IP, or according to being registered as an IP. If we would have made these subgroups, we then probably could have anticipated better on the IPs' individual needs. However, the numbers of participating IPs in our study were too small for such an approach. Concerning the fifth and sixth elements, by performing a needs assessment and by consulting IPs in semi-structured interviews we managed to develop and adapt the implementation strategy. For instance, the IPs asked for a tailor-made training and for tools to facilitate them applying the guidelines. As to the seventh element; we used a mix of methods tailored to needs and preventing obstacles, albeit only at the level of the IPs, counting with the costs and our limited budget.

**Table 9.1:** Elements of effective implementation [18].

Element	Sign
1. A systemic approach to and good planning of implementation activities is needed most of the time.	+
2. Focus on the innovation – is it a good product?	+
3. Diagnostic analysis of the target group and setting should take place before the start of the implementation.	+
4. Subgroups within the target group may be at different stages of the change process and have different needs – segmentation within the target group should be allowed for.	-
5. The target group should be involved in the development and adaptation of the innovation, as well as in planning the implementation.	+
6. The choice of implementation activities should link with the results of the diagnostic analysis.	+
7. Usually, a single method or measure is insufficient – search for cost-effective mix of methods tailored to the identified obstacles and incentives to change.	+
8. Make a distinction between the phases of implementation (dissemination, implementation and integration) – different measures and strategies are effective at different stages.	-
9. Take the appropriate measures for each of the various levels – national, local, team, practice and professional.	-
10. Continuous evaluation of both the implementation process and its results required.	-
11. Make implementation an integral part of the existing structures.	-

The last four negative scores in this box can be attributed to the fact that our implementation strategy was evaluated in an efficacy study, and not in an effectiveness study. Therefore, the elements with negative scores imply recommendations for implementing the strategy in real practice. The guidelines for depression were already disseminated and implemented at the Institute approximately one year before this research project started. In this research project the efficacy of an alternative specific implementation strategy was studied in an experimental setting, leaving no room for targeting the strategy at the organizational level of the Institute. For instance, when implementing the developed strategy at the Institute various levels should be involved: the IPs at the front offices, the staff IPs, managers, the board, and the National Association of Insurance Physicians. Only then the strategy would become an integral part of the structure at the Institute, and continuous evaluation would be possible.

### Barriers for implementation of guidelines

'Why don't physicians follow clinical guidelines' is the title of an article, that can be regarded as characteristic for literature concerning guideline adherence [19]. A lot of research has been done in order to find and study barriers to successful implementation of guidelines [20]. The findings of all these studies make clear that implementing guidelines is a complex process, depending on multiple factors at different levels. Somewhere in that process a weak link often occurs, explaining the gap between evidence-based medicine of the guidelines and which part ends up in the physicians consultation room or what reaches the patient. Well known barriers to effective implementation on the level the physicians are: negative attitudes, beliefs, e.g. low outcome expectancies, values and individual perceptions that can hinder motivation for change [21]. Examples are: lack of time, lack of knowledge, and lack of skills. We evaluated the experiences of the IPs with our implementation strategy, and some of them mentioned available time and changing work routines as possible barriers. Actually, compared to other physicians, IPs might have a kind of familiarity with guidelines, because their work always takes place within the framework of social benefits legislation.

At organizational level barriers might occur, such as organizational structures and differences in policies at front offices, problems of integrating guidelines in the culture of work processes, lack of collaboration between providers of guidelines and management, lack of facilitations for training, lack of communication with physicians, and lack of measures for monitoring and evaluating the implementation [21]. In this study we did not investigate any barriers for the implementation strategy at level of the organization (Institute). Potential barriers at the Institute might be: apart from time and costs, the fact that the guidelines for depression already have been implemented before and that the impact of the newly developed implementation strategy in real practice on outcomes such as inter-IP agreement or inflow into the benefits remain unknown. Hence, it might be difficult to convince the Institute stakeholders of the surplus value of the developed implementation strategy. A good opportunity could be the moment when the guidelines for depression need to be updated, e.g. when the DSM V is published. Implemented knowledge is not static. Knowledge may be better regarded as a semi-finished product that gets adjusted in practice to make it work [22]. Another opportunity would be to share our experience with the implementation of the guidelines for depression, with future implementations of other insurance medicine guidelines.

### Considerations of insurance physicians' behaviour regarding guideline implementation

At the start of this research project, we chose for the ASE model because this model appeared to be useful for describing behaviour and changes in that behaviour in relation to interventions in the field of occupational health [3, 13]. These studies had outcomes on patient level, e.g. return to work. Our study, however, had only outcomes on the level of the physician, e.g. guideline adherence.



Unfortunately, our findings presented in Chapter 4 showed that the ASE model, in a cross-sectional analysis on baseline data, could not be confirmed at its main causal path leading from intention to behaviour. In our study population, we did not find a positive relation between IPs' intention to use the guidelines and behaviour, in this case using the guidelines. However, the model appeared useful for exploring the IPs' behaviour towards guidelines in the phase preliminary to our implementation strategy. A remarkable finding was that the IPs reported to be influenced by their colleagues and the way guidelines were implemented. Obviously, to IPs it does matter the way guidelines are being implemented, and how their colleagues think of using guidelines. In the development of our implementation strategy, we could anticipate to these findings by suiting to the IPs' needs.

After having received the implementation strategy the IPs' behaviour towards the guidelines for depression changed positively indeed (Chapter 6). All measured ASE determinants of IPs' behaviour e.g. attitude, self-efficacy, knowledge and skills, and intention changed significantly in the expected direction. That left us with the question whether these changes in ASE-determinants could be linked to observed behaviour, guideline adherence. In Chapter 6 it was shown that changes in only one ASE-determinant, knowledge and skills, was weakly positively related to improvement in guideline adherence. It seems unlikely that the implementation strategy only had influence on IPs' knowledge and skills. Probably, the implementation strategy influenced IP's guideline adherence in other ways than described by the ASE model. For instance, the implementation strategy could have had an influence on guideline adherence, by influencing the IPs' motivation to use guidelines. The ASE model does not include motivation. However, motivation plays a central role in the MODE model [23]. This MODE model postulates that attitude can guide behaviour in a spontaneous manner. Attitude may be activated from memory automatically upon the individuals' encountering the attitude object. Instead of making well-considered choices, IPs might have a direct and strong automatically activated attitude regarding guidelines, because of their experiences with the implementation of 20 guidelines in the recent past. According to the MODE model motivation goals can have a strong influence on that automatically activated attitude. At the moment that IPs are confronted with another implementation strategy of guidelines, the IPs' automatic attitude could be moderated or mediated by their motivation and in this way could influence their guideline adherence. Measuring the IPs' motivation and automatically activated attitude towards guidelines might have given more insight in IP's behaviour, than using only the ASE model.

Furthermore, the IPs in our study selected themselves for participation and probably they already might have had a certain level of motivation for learning to apply guidelines, or they might have selected themselves for reasons of gaining accreditation points needed for their registration as an IP. The IPs who selected themselves might already use guidelines in practice or might have been using guidelines in the recent past. Another

useful model could then be a model that predicts behaviour in the future by behaviour in the past [24]. The behaviour that someone has shown in the past is kept in memory, and that behaviour automatically repeats at a certain moment without interference of attitudes. People who have behaved in a certain way at one point in time are likely to do so again [25]. IPs might be creatures of habit as well. Some of the IPs who participated in our study mentioned changing their work routines needed for applying the guidelines for depression as a potential barrier to the implementation strategy.

## Practical implications of the implementation strategy

### For IPs

Assessing work disability of clients with depression is a challenging task for the IP. The guidelines, when applied carefully, could support the IPs in fulfilling this task. Looking back to the case history from the General Introduction, an IP can make clear and well argued decisions based on the guidelines. For instance, assessing the diagnosis by the DSM criteria makes clear whether a client has the diagnosis depression or not by asking for and counting of the symptoms. Furthermore, by counting and defining the severity of the symptoms, the severity of the depression can be assessed. According to the guidelines, the severity of the depression has implications for the work limitations. In this way the work limitations can be assessed and reported in a well argued way. For instance, if the teacher from the case history in the General Introduction, has serious sleeping problems as one of the symptoms, he might find difficulties in keeping up his attention and concentration to the required level that is needed for teaching all day. One of the practical problems for IPs, for example, is that they have to decide whether the client has limitations in the number of working hours [26, 27]. Therefore, we facilitated the IPs with a tool, the summary desk mat, showing besides DSM-IV diagnostics an application of the ICF-model to the guidelines for depression that covers the potential work limitations of this disorder. By exercising assessments of realistic cases and receiving feedback on their performances by trainer IPs, the IPs learn, facilitated by the tools, how to deal with that kind of practical problems. This is an example of an item where the developed implementation strategy meets the IPs' needs. The summary desk mat and the checklists support the IP to follow the guidelines, and stimulate the IP to work systematically and more analytically. Such a work style might differ from their usual working routines, in which IPs might show the tendency to follow their clinical judgement, for instance based on recognition of patterns. This might give the impression of jumping to early conclusions, or they might stick to the findings and conclusions of the physicians who have assessed the same client before. The IPs who received the implementation strategy expected to perform better assessments of clients with depression, but at the same time they perceived changing their working routines in

order to apply the guidelines as a potential barrier (Chapter 8). By following guidelines it can be argued that IPs can improve their work disability assessment reports, without loss of their professional autonomy, at the costs of changing their routines. That should be worth trying.

### Implications for IPs

- This implementation strategy provides the translation of evidence-based medicine from the guidelines into practice
- Applying guidelines for depression can improve the quality of IPs' assessment reports
- Applying guidelines for depression makes the IPs' work more transparent to other IPs
- Applying the guidelines for depression can support IPs in the difficult task of translating disorders to functional abilities

### For stakeholders and policymakers

When the insurance medicine guidelines were developed by the Dutch Health Council, the job was not finished yet. The implementation of the guidelines came separately and afterwards, carried out by the Institute. Developing guidelines together with an implementation strategy, and then implementing the guidelines by the same organization should be recommended. We recommend that this developed implementation strategy should be carried out in practice at the Institute and then evaluated. This implementation strategy provides the opportunity for monitoring physicians' performances in applying the guidelines for depression. That could be possible for other guidelines as well, because the majority of the 20 implemented guidelines can be modelled by the same kind of PIs as those of the guidelines for depression.

One practical barrier for use in practice of the PIs might be that at least two judges are needed for sufficient reliability, and that the judges have to be trained in applying the PIs. In case of disagreement between the two judges, a consensus procedure with a third independent judge is needed. That is rather time-consuming and inefficient.

Managing IPs' quality of work contributes to the aim of the Institute for being a national centre of insurance medicine expertise. This thesis does not include a cost-effectiveness study. Therefore, we only can say that by using the implementation strategy, the Institute invests in quality of disability assessment reports. To what costs is unknown. For promoting the implementation strategy, a proven cost-effectiveness of the strategy could have been useful. On the level of national societal benefits it is also of interest to know, whether the implementation of guidelines in general has impact on the total inflow in and costs of disability benefits. Applying guidelines for depression improves inter-IP-agreement in the scoring of the List of Functional Abilities.



At the Institute, compared to e.g. general practitioners or occupational physicians, conditions for implementing guidelines at the level of organization seem to be good. Nearly all IPs working at the Institute are employees of the Institute, which makes it easy to manage them as a group. For instance, all IPs are supervised by a staff IP, who is responsible for the quality of IPs' work. Resuming, implementation of insurance medicine guidelines at the Institute might work out successfully.

### Implications for Institute stakeholders

- The developed implementation strategy for the insurance medicine guidelines for depression could be evaluated in a pilot effectiveness study. Barriers to and facilitators of this implementation strategy should firstly be identified in real practice, before enrolment throughout the Netherlands
- Continuous monitoring of IPs' performances on applying the insurance medicine guidelines for depression has been made possible
- The developed implementation strategy for the insurance medicine guidelines for depression can be adapted to other insurance medicine guidelines
- The impact of nationwide implementation of the insurance medicine guidelines for depression on disability benefits due to depression remains unknown and should be investigated

### For medical education programmes aiming at IPs

Once they are in, IPs do appreciate the implementation strategy. At the start, however, small numbers of IPs were willing to participate in the educational programme (i.e. the tailor-made training). Therefore, more attention should be paid to the recruitment of IPs to attract them. The IPs' interest has to be roused by offering an attractive educational programme, which can be provided by multifaceted programmes such as our developed implementation strategy.

Such an educational programme has to meet the IPs' needs, and should contain a balanced mixture of evidence-based medicine training and practical exercising [28]. Realistic case histories are essential for practising the application of guidelines in an educational programmes. The next step then, is to apply guidelines in real practice. Unfortunately, regarding the results of our study, we know that the developed implementation strategy as an entity works, but it remains unclear which specific part of it caused the differences.

### Implications for educational programmes aiming at IPs

- Participation of IPs in educational programmes on voluntary base might result in a low reach. Attention should be paid to the recruitment procedures of IPs for educational programmes
- IPs are satisfied when receiving a multifaceted educational programme



- IPs appreciate practical tools to facilitate them in applying guidelines for depression in practice
- IPs appreciate being trained by IP-trainers

## Recommendations for further research

Firstly, research should be focused on the quality of the guidelines themselves. It would be interesting to know, whether the insurance medicine guidelines of depression actually meet the standards of AGREE [29]. In an earlier study on guidelines and the evaluation of work disability it was found that the Dutch guidelines on average scored low at rigour of development and applicability, according to the standards of AGREE [30]. Secondly, a logical successor to this thesis would be an effectiveness study of the guidelines for depression, carried out at the Institute. In that case, long-term effects of the implementation strategy should be measured and evaluated, extended with outcomes at client level, which this thesis lacks. Another idea is to study the effectiveness of similarly implemented guidelines for different diseases in a before-after design. Thirdly, the quality of various educational interventions in the field of insurance medicine should be investigated in a review of the literature. Beside the educational intervention of the present study, this review could include for instance the evidence-based medicine course [31], or different training programmes for interviewing the client such as: “The communication skills training” [32] or “The disability assessment structured interview” [33]. Subsequently, all these interventions with proven quality could be used for upgrading the post-graduate course for being registered as an IP. Finally, it appears from this study (see Chapter 7) and other studies among Dutch IPs that trained IPs tend to assess more work limitations compared to untrained colleagues [26, 27]. When an IP gathers more information on a client, e.g. due to adherence to guidelines or after following other ways of structured assessments such as “The disability structured interview”, the number and the severity of work limitations seem to increase, which may possibly lead to an increase in disability benefits. This last point makes also clear that more research is needed on one of the main tasks of IPs, which is the translation from diseases and disorders to people’s work limitations or functional abilities [30]. At the level of disability benefits the question remains whether IPs tend to fail in recognizing diseases and the accompanying work limitations, i.e. IPs tend to under-diagnose leading to fewer registrations of disease burden. One might also argue that effects of medicalization become visible due to guidelines, i.e. when following guidelines, IPs tend to over-diagnose leading to more registrations of disease and subsequently to more inflow into the disability benefits.

## Recommendations for future research at the Institute

- To carry out an effectiveness study of the developed implementation strategy for the insurance medicine guidelines for depression
- To study the effectiveness of recently implemented guidelines, such as chronic renal failure or diabetes, in a before-after design
- To study the process of work disability assessment by IPs in general, with the focus on the translation from disease or disorder to work limitations or functional abilities
- To study the influence of guidelines on the assessment of work limitations and disability benefits

## Conclusions

Recently, many insurance medicine guidelines have been developed for supporting the work disability assessments carried out by IPs. There was no existing evidence or experience with the implementation of insurance medicine guidelines. Therefore, we developed an implementation strategy for one of these guidelines, i.e. the one on depression. This implementation strategy, which was characterized by a multifaceted approach and meeting the needs of the IPs and stakeholders contributed to the guideline adherence of the IPs in a controlled setting. Further research to the effectiveness of the developed implementation strategy in real practice is recommended. Here are the conclusions of this thesis:

- The developed implementation strategy for the guidelines for depression improves the guideline adherence of IPs and their knowledge of the insurance medicine guidelines
- The psychological behavioural model (ASE model) gave support to describe IPs’ behaviour towards guidelines, but could only partly be confirmed
- IPs’ determinants of behaviour with regard to the guidelines for depression, such as attitude, self-efficacy, knowledge and skills, and intention to use the guidelines, changed in a positive way as a result of the developed implementation strategy
- Changes in IPs’ knowledge and skills due to the implementation strategy were weakly positively related to their improvements in guideline adherence
- The number and severity of work limitations of a disability assessment tend to increase when applying the insurance medicine guidelines for depression
- The inter-IP agreement of the assessment of work limitations increased due to the developed implementation strategy
- Time needed for applying the insurance medicine guidelines and changing work routine for using the guidelines in practice were perceived as barriers to the implementation strategy by a part of the IPs
- The IPs positively evaluated the developed implementation strategy and they expected to improve their assessments of clients with depression

## References

1. Bartholomew LK, Parcel GS, Kok GJ: Intervention mapping: a process for developing theory- and evidence-based health education programmes. *Health Educ Behav* 1998, **25**:545-63.
2. Vermeulen S, Anema J, Schellart A, Van Mechelen W, Van der Beek A: Intervention mapping for development of a participatory return-to-work intervention for temporary agency workers and unemployed workers sick-listed due to musculoskeletal disorders. *BMC Public Health* 2009, **9**:216.
3. Van Oostrom SH, Anema JR, Terluin B, Venema A, De Vet HCW, Van Mechelen W: Development of a workplace intervention for sick-listed employees with stress-related mental disorders: Intervention Mapping as a useful tool. *BMC Health Serv Res* 2007, **7**:127.
4. Grol R: Improving the quality of medical care: building bridges among professional pride, payer profit, and patient satisfaction. *JAMA* 2001, **286**:2578-85.
5. De Leng B, Dolmans D, Van de Wiel M, Muijtjens A, Van der Vleuten C: How video cases should be used as authentic stimuli in problem-based medical education. *Med Educ* 2007, **41**:181-8.
6. Heath C, Luff P, Svensson MS: Video and qualitative research: analysing medical practice and interaction. *Med Educ* 2007, **41**:109-16.
7. Dequeker J, Jaspert R: Teaching problem-solving and clinical reasoning: 20 years experience with video-supported small-group learning. *Med Educ* 1998, **32**:384-9.
8. Hatala R, Issenberg SB, Kassen B, Cole G, Bacchus CM, Scalese RJ: Assessing cardiac physical examination skills using simulation technology and real patients: a comparison study. *Med Educ* 2008, **42**:628-36.
9. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG: CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Int J Surg* 2012, **10**:28-55.
10. Praetorius F: Guidelines and autonomy of physician decision. A critical survey under ethical aspects. *Z Arztl Fortbild Qualitatssich* 2005, **99**:15-23.
11. De Vries H, Dijkstra M, Kuhlman P: Self-efficacy: third factor besides attitude and subjective norm as a predictor of behavioral intention. *Health Educ Res* 1988, **3**:273-82.
12. Ajzen I: The theory of planned behavior. *Organ Behavi and Hum Decis Process* 1991, **50**:179-211.
13. Rebergen D, Hoenen J, Heinemans A, Bruinvels D, Bakker A, Van Mechelen W: Adherence to mental health guidelines by Dutch occupational physicians. *Occup Med* 2006, **56**:461-8.
14. Van Rijssen HJ, Schellart AJ, Anema JR, Van der Beek AJ: Determinants of physicians' communication behaviour in disability assessments. *Disabil Rehabil* 2011, **33**:1157-68.
15. Perkins MB, Jensen PS, Jaccard J, Gollwitzer P, Oettingen G, Pappadopulos E, Hoagwood KE: Applying theory-driven approaches to understanding and modifying clinicians' behavior: what do we know? *Psychiatr Serv* 2007, **58**:342-8.
16. Godin G, Bélanger-Gravel A, Eccles M, Grimshaw J: Healthcare professionals' intentions and behaviours: a systematic review of studies based on social cognitive theories. *Implement Sci* 2008, **3**:36.
17. Hanbury A, Wallace L, Clark M: Use of a time series design to test effectiveness of a theory-based intervention targeting adherence of health professionals to a clinical guideline. *Br J Health Psychol* 2009, **14**:505-18.
18. Grol R, Wensing M, Eccles MP: *Improving Patient Care: the implementation of change in clinical practice*. Oxford: Elsevier, 2005.
19. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al.: Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 1999, **282**:1458-65.



20. Grol R: Personal paper. Beliefs and evidence in changing clinical practice. *BMJ* 1997, **315**:418-21.
21. Addington D, Kyle T, Desai S, Wang J: Facilitators and barriers to implementing quality measurement in primary mental health care: systematic review. *Can Fam Physician* 2010, **56**:1322-31.
22. Wensing M, Bal R, Friele R: Knowledge implementation in healthcare practice: a view from The Netherlands. *BMJ Qual Saf* 2012, **21**:439-42.
23. Fazio RH: Multiple processes by which attitudes guide behavior: the MODE model as an integrative framework. In: M. P. Zanna (Ed.), *Advances in experimental social psychology* 1990, **23**:75-109. New York: Academic Press.
24. Fredericks AJ, Dosett L: A comparison of the Fishbein-Ajzen and the Bentler-Speckart Models. *Journal of Personality and Social Psychology* 1983, **45**:501-12.
25. Albarracín D, Wyer RS: The cognitive impact of past behaviour: influences on beliefs, attitudes, and future behavioural decisions. *Journal of Personality and Social Psychology* 2000, **79**:5-22.
26. Spanjer J, Krol B, Brouwer S, Groothoff JW: Sources of variation in work disability assessment. *Work* 2010, **37**:405-11.
27. Schellart AJ, Mulders H, Steenbeek R, Anema JR, Kroneman H, Besseling J: Inter-doctor variations in the assessment of functional incapacities by insurance physicians. *BMC Public Health* 2011, **11**:864.
28. Berkhof M, Van Rijssen HJ, Schellart AJ, Anema JR, Van der Beek AJ: Effective training strategies for teaching communication skills to physicians: an overview of systematic reviews. *Patient Educ Couns* 2011, **84**:152-62.
29. 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.
30. De Boer WE, Bruinvels DJ, Rijkenberg AM, Donceel P, Anema JR: Evidence-based guidelines in the evaluation of work disability: an international survey and a comparison of quality of development. *Int Arch Occup Environ Health* 2010, **83**:201-8.
31. Kok R, Hoving JL, Verbeek JH, Schaafsma FG, Smits PB, Van Dijk FJ: Evaluation of a workshop on evidence-based medicine for social insurance physicians. *Occup Med* 2008, **58**:83-7.
32. Van Rijssen HJ, Schellart AJ, Anema JR, De Boer WE, Van der Beek AJ: Systematic development of a communication skills training course for physicians performing work disability assessments: from evidence to practice. *BMC Med Educ* 2011, **11**:28.
33. Spanjer J, Krol B, Brouwer S, Popping R, Groothoff JW, Van der Klink JJ: Reliability and validity of the Disability Assessment Structured Interview (DASI): a tool for assessing functional limitations in claimants. *J Occup Rehabil* 2010, **20**:33-40.

# **Summary**

**Development and evaluation of  
an implementation strategy for  
insurance medicine guidelines  
for depression**



## General introduction

Since 2006, 20 Dutch insurance medicine guidelines have been developed for various diseases. These guidelines were steadily implemented by the Dutch Institute for Employee Benefits Schemes (UWV) in insurance physician's (IP) practice. The implementation of all these guidelines was never evaluated, and therefore, the Knowledge Center for Insurance Medicine (KCVG) decided to start a research project on one of the insurance medicine guidelines. The guidelines for depression were chosen, because of their relevance for society. Depression causes a sizeable part in the total amount of work disability world wide, and in the Netherlands depression as diagnosis takes the first place for inflow into the disability benefits. Insurance medicine in practice at the UWV was explained with regard to the context of insurance medicine guidelines. A case history of a client with depression was introduced to make clear that the assessment of a client with depression is not an easy task for an IP, and that the use of guidelines not only might relieve this task, but also might have effect on the assessment of work limitations of a client by an IP. IPs and various other stakeholders within the UWV had several questions concerning the implementation of guidelines:

- Which strategy can be developed to implement the guidelines for depression, in order to promote use by IPs?
- What are the effects of such a developed implementation strategy on the guideline adherence of the IPs and on their knowledge of the guidelines?
- What are the effects of such a developed implementation strategy on the behavioural determinants of the IPs regarding the use of the guidelines?
- What are the effects of such a developed implementation strategy on the number and severity of work limitations when applying the guidelines?
- What are the effects of such a developed implementation strategy on the inter-IP agreement in the work disability assessments of the IPs?
- What are the effects of such a developed implementation strategy on the satisfaction of the IPs?

## Development of the implementation strategy

Chapter 2 describes the development of the implementation strategy, using the Intervention Mapping (IM) method. This IM method supplies a stepwise framework, to develop and evaluate an intervention. The development of the implementation strategy was supported by a psychological behavioural model, the Attitude, Social influence,

self-Efficacy model (ASE model). By interviewing IPs in practice and stakeholders, we analysed the context at the UWV. Additionally, we performed a needs assessment of the IPs in practice regarding implementation of guidelines. Finally, we designed the implementation strategy after consulting literature, IP trainers and various experts in the field of guideline implementation. IM provided the planning tool for mapping the path of the intervention development from a needs assessment to the potential solution. In this way, our approach differed from the usual implementation of guidelines at the UWV, which was merely a top-down approach. Intervention mapping appeared to be a useful but time-consuming method for the development of a multifaceted implementation strategy for the guidelines for depression. The developed implementation strategy consisted of a multifaceted training, in which the IPs, facilitated with various tools, should learn to apply the guidelines for depression. The IPs should be trained by two trainer IPs in interactive subgroups and should receive feedback on their performances. The evidence-based theory of the guidelines was translated for use in practice, and summarized on a desk mat.

## Development and reliability of performance indicators

Chapter 3 presents the development of performance indicators (PI) and their reliability. For the evaluation of the implementation strategy, we had to develop an instrument to measure the IPs' guideline adherence. PIs for measuring guideline adherence were developed with the help of experts. We ended up with six PIs in the form of decision trees, reflecting the most important elements of the guidelines. The PIs indicate whether or not an assessment report is adequate according to the guidelines. With these PIs, the IPs' guideline adherence in the disability assessment reports as a whole could be measured in a sum score.

Eight selected IPs (Test IPs) were trained in applying the PIs. After the training they applied the PIs on ten constructed disability reports of clients with depression, to test the reliability of the application of the PIs. The Test IPs considered the PIs as a content valid and feasible instrument. The PIs were found to be a reliable instrument (ICC 0.70 or higher) if at least two Test IPs were involved.

## Behaviour of the insurance physicians

Chapter 4 describes the explorations of the determinants of the IPs' behaviour towards guidelines in general, and more specific to the guidelines for depression. As a starting point for the study of the IPs' behaviour we used the ASE model. We developed questionnaires for measuring the baseline characteristics of the participating IPs, the ASE determinants of the IPs' behaviour towards guidelines, and the interfering factors in the model. The majority (85%) of the participating IPs reported to use at least some elements of the guidelines for depression. We studied the IPs' behaviour towards the guidelines for depression by analysing the data of the questionnaires with the use of



structural equations modelling. It appeared that the IPs' intentions to use the guidelines for depression and their self-reported use of these guidelines were related to the influence of colleagues, their self-efficacy, and the way the guidelines are implemented. However, the ASE model could only partly be confirmed, because we did not find a relationship between intention and self-reported use.

### **The evaluation of the implementation strategy**

Chapter 5 presents the evaluation of the implementation strategy. The main aim of this study was to evaluate whether the implementation strategy would improve the guideline adherence of the IPs. The secondary outcome of this study was the IPs' knowledge of the guidelines for depression. The developed strategy was evaluated in an experiment in a controlled setting. We compared the developed implementation strategy (intervention group) to the usual methods of implementing guidelines at the UWV (a 'placebo' training for the control group) by measuring their performances in disability assessments of clients with depression in the experiment. All participating IPs had to assess the disability of four different clients with depression, played by actors, and presented at video. The IPs wrote two disability assessment reports before, and another two after the implementation strategy. The guideline adherence in the disability reports of the participating IPs was assessed by trained Test IPs using the PIs (see Chapter 3).

The IPs who received the implementation strategy performed significantly better on the PIs (on average 4.44 on the theoretically mean PI sum score 1.00-5.67) than the IPs from the control group (on average 3.32). Higher scores on the PI sum score on a report indicated that the report was more in concordance with the guidelines. The IPs knowledge of the guidelines was separately tested, and the IPs in the intervention group who had received the implementation strategy performed better on the knowledge test than the IPs from the control group. We concluded that the developed implementation strategy for the guidelines for depression improved the guideline adherence of IPs and their knowledge of the guidelines in an experimental, controlled setting.

### **The changes in determinants of insurance physicians' behaviour after the implementation strategy**

Chapter 6 describes the changes in the behavioural determinants of the IPs towards the guidelines for depression caused by the implementation strategy. These behavioural determinants were measured using questionnaires developed on base of the ASE model, before and three months after the training in applying the guidelines for depression. The IPs' behavioural determinants, based on the ASE model, changed positively after having received the implementation strategy compared to the control group. All investigated determinants of the ASE model (i.e. attitude, self-efficacy, knowledge and skills, and the intention to use the guidelines) changed significantly when the intervention group was compared to the control group. After the implementation strategy, attitude and intention

to use the guidelines improved with 12%, self-efficacy with 10%, and knowledge and skills with 5%. Only changes in self-reported knowledge and skills were related to the improvements in observed guideline adherence of the IPs, as measured with the PIs. However, this relation was only weak.

### **Number and severity of work limitations and inter-rater reliability of disability assessments when applying insurance medicine guidelines for depression**

The aim in Chapter 7 was to study the influence of the implementation strategy on the IPs' assessment of the work ability of clients with depression, using the standardised form of the List of Functional Abilities (LFA). The IPs who participated in the controlled experiment assessed the work disability of four different clients with depression. They scored for each client the accompanying LFA. After the implementation strategy: 1) IPs applied significantly more numerous and severe work limitations, and 2) the inter-rater reliability of work disability assessments of clients with depression by IPs was higher in the intervention group than in the control group, the latter was, however not significant.

IPs should be aware of the fact that following guidelines often implies collecting more information about a client, which can lead to the finding of more work limitations, both in number and in severity, in that client. For policy makers it is interesting to know that it is possible to improve uniformity in the work disability assessments by proper training IPs in applying the guidelines.

### **The process evaluation of the implementation strategy**

Chapter 8 presents how the implementation strategy was perceived by the IPs. The 42 IPs who participated in our study were highly satisfied with the implementation strategy overall (a mean score of 7.7 on a 1-10 scale). In particular, they appreciated the training and the summary desk mat. Immediately after having received the implementation strategy, the majority (81%) of the IPs expected to improve their assessments of clients with depression by applying the guidelines, and 86% expected to improve their assessment reports. After three months 96% of the IPs considered the implementation strategy as having been useful. Time needed for applying the guidelines was mentioned by some of the IPs as a barrier, and some of them had concerns about changing their work routines regarding applying the guidelines. A weakness of our study was that we reached only 5% of the IPs working at the UWV.

### **The general discussion**

Chapter 9 presents the answers to the questions asked in the General Introduction as the main findings of this thesis. In addition, methodological considerations and recommendations for further research were addressed.



The use of the Intervention Mapping method, for the development, the planning, and finding the right approach of the implementation strategy was described. The design of the study was discussed by mentioning the pros and cons of the efficacy design versus the effectiveness design with regard to the context of the UWV. The original plan, to evaluate the developed strategy in an effectiveness study at the UWV, failed because of practical reasons. Therefore, we carried out an efficacy study for the implementation strategy in the form of an experiment in a controlled setting. The primary outcome of the research project, guideline adherence with the accompanying measurement instrument was clarified. Educational aspects of the implementation strategy were reflected on in a broader perspective. The exploration of IPs' behaviour towards guidelines by the use of psychological models was discussed with regard to literature. In addition, the developed implementation strategy was compared to the features of effective implementations described in literature. In general, barriers to the implementation of guidelines are an important issue, and therefore we added a paragraph on the potential barriers for our implementation strategy.

### **Conclusions and implications**

An implementation strategy for the insurance medicine guidelines for depression was developed and evaluated. The efficacy of developed implementation strategy was demonstrated. In this study it was shown that IPs who received the implementation strategy adhered better to the guidelines for depression, and had a better knowledge of these guidelines than a control group who had received the usual implementation by the UWV. Furthermore, the IPs appreciated the implementation strategy, as was also confirmed by a positive change in their behaviour towards the guidelines for depression. Finally, the inter-IP agreement improved, indicating more uniformity between the IPs in their work disability assessments of clients with depression.

The overall conclusion is that we successfully managed to develop a multifaceted implementation strategy for the guidelines for depression. However, evaluation of this implementation strategy in real practice remains still needed. Besides, the implementation strategy was mainly evaluated on the level of the IPs, and in a specific controlled setting. Evaluations of other levels, such as the organization or the client are also important. The results of this thesis have various practical implications for IPs, stakeholders, and for medical education programs aiming at IPs. For IPs: this implementation strategy can improve the quality of the IPs work disability reports, which makes the IPs' work more transparent to others, and uniformity of the assessments can be enhanced. For UWV stakeholders: this implementation strategy for the guidelines for depression should be applied throughout the Netherlands, and can also be adapted to other insurance medicine guidelines. Monitoring IPs' performances on applying guidelines has been made possible. For educational programmes aiming at IPs: the translation of evidence-based medicine from the guidelines to IPs' practice can be achieved by

using experienced IP trainers and with help of realistic case histories of UWV clients. The efficacy of the developed implementation strategy for the guidelines for depression has been demonstrated. This implementation strategy contributes to quality improvement in insurance medicine.

# **Samenvatting**

**Ontwikkeling en evaluatie  
van een implementatie-  
strategie voor het verzekerings-  
geneeskundig protocol  
Depressieve stoornis**



## Algemene inleiding

Sinds 2006 zijn er 20 verzekeringsgeneeskundige protocollen ontwikkeld voor verschillende ziektebeelden in Nederland. Deze protocollen werden op voortvarende wijze door het UWV (Uitvoeringsinstituut Werknemers Verzekeringen) geïmplementeerd in de verzekeringsgeneeskundige praktijk. De implementatie van deze protocollen was nog niet geëvalueerd, en daarom besloot het Kenniscentrum Verzekeringsgeneeskunde (KCVG) een onderzoeksproject te beginnen met als onderwerp één van deze protocollen. De keuze viel op het protocol Depressieve stoornis vanwege de maatschappelijke relevantie. Een aanzienlijk deel van het totale volume van arbeidsongeschikten wereldwijd wordt veroorzaakt door depressie. In Nederland neemt depressie als diagnose de eerste plaats in bij de instroom in de arbeidsongeschiktheid WIA (Wet Inkomensvoorziening Arbeidsongeschikten). De praktijk bij het UWV met betrekking tot de verzekeringsgeneeskundige protocollen wordt toegelicht aan de hand van een voorbeeld uit de praktijk. Deze casus maakt duidelijk dat de arbeidsongeschiktheidsbeoordeling van een cliënt met een depressieve stoornis geen eenvoudige taak is voor een verzekeringsarts (VA). Het toepassen van een protocol zou deze taak niet alleen kunnen verlichten, maar zou ook gevolgen kunnen hebben voor de uitkomst van een arbeidsongeschiktheidbeoordeling, namelijk het aantal en de ernst van de beperkingen die de verzekeringarts bij een cliënt vaststelt. VA'n en beleidsmakers van het UWV hadden verscheidene vragen met betrekking tot de implementatie van de verzekeringsgeneeskundige protocollen:

- Welke strategie kan er worden ontwikkeld met het doel het gebruik van het protocol Depressieve stoornis door verzekeringsartsen te promoten?
- Wat zijn de effecten van zo'n implementatiestrategie in een gecontroleerde omgeving op:
  - De gedragsdeterminanten van de VA'n met betrekking tot het gebruik van het protocol Depressieve stoornis
  - Het volgen van het protocol Depressieve stoornis door de VA'n en hun kennis van dat protocol
  - Het aantal en de ernst van de beperkingen bij toepassing van het protocol Depressieve stoornis
  - De interdoktersvariatie in de arbeidsongeschiktheidsbeoordelingen van de VA'n
  - De tevredenheid van de VA'n met het toepassen van het protocol Depressieve stoornis

## Ontwikkeling van de implementatiestrategie

In hoofdstuk 2 wordt de ontwikkeling van de implementatiestrategie beschreven. Hiervoor werd de 'Intervention Mapping' (IM) methode gebruikt. Deze IM methode voorziet in een gestructureerde en stapsgewijze aanpak voor de ontwikkeling en de evaluatie van de strategie, ondersteund door theorie en gericht op de praktijk.

Voor de theorie gebruikten wij het 'Attitude, Social influence, self-Efficacy' model (ASE-model). Om aan te kunnen sluiten bij de praktijk interviewden wij VA'n en beleidsmakers uit de praktijk van het UWV. Bovendien voerden we een inventarisatie uit van de behoeftes van de VA'n met betrekking tot de implementatie van protocollen. Ten slotte ontwierpen we de implementatiestrategie op basis van deze interviews, beschikbare literatuur, en na consultatie van VA-trainers en diverse experts met kennis van protocol- of richtlijn-implementaties. Voor onze implementatiestrategie voldeed IM als planningshulpmiddel vanaf de inventarisatie van de behoeftes van de VA'n tot aan het eindproduct, de ontwikkelde implementatiestrategie. Hierin verschilde onze implementatiestrategie met die van het UWV, die meer gekenmerkt werd door een 'top-down' benadering. De IM- methode nam echter wel veel tijd in beslag. De ontwikkelde implementatiestrategie bestond uit een rijk geschakeerde training, waarin de VA'n, ondersteund door verscheidene hulpmiddelen, leerden het protocol Depressieve stoornis toe te passen. De 'evidence-based medicine' van het protocol Depressieve stoornis werd vertaald naar de praktijk van de VA met behulp van o.a. een bureaulegger, die een handige samenvatting bevatte van het protocol. De VA'n, verdeeld in subgroepen, werden interactief getraind door twee VA-trainers. De VA'n kregen op groepsniveau feedback op hun prestaties.

## Ontwikkeling en betrouwbaarheid van de performance indicatoren

Aangezien er geen meetinstrument bestond voor de mate waarin VA'n een verzekeringsgeneeskundig protocol volgen (guideline adherence) hebben we dat meetinstrument moeten ontwikkelen. In Hoofdstuk 3 wordt de ontwikkeling en de betrouwbaarheid van het meetinstrument (performance indicatoren, (PI'n)) beschreven. Daarbij hebben we de hulp ingeroepen van deskundigen met ervaring in het toetsen van VA-rapportages, zoals senior verzekeringsartsen met een staffunctie, die samen met de onderzoekers de performance indicatoren ontwikkelden. Het resultaat was zes performance indicatoren in de vorm van beslisbomen, elk gebaseerd op een hoofd element van het protocol Depressieve stoornis. Met deze PI'n kan worden aangegeven of een VA-rapportage adequaat of niet-adequaat is op het desbetreffende onderwerp van de PI. Met behulp van deze PI'n kan de 'guideline adherence' van een gehele VA-rapportage worden gemeten als een somscore. Acht geselecteerde VA'n werden getraind in het toepassen van de PI'n. Na deze training pasten deze acht VA'n de PI'n toe op tien geconstrueerde VA-rapportages van cliënten met een depressie. Op deze wijze werd de betrouwbaarheid van de PI'n getest. Deze zogenaamde Test-VA'n beschouwden de PI'n als een op inhoud valide en bruikbaar instrument. De PI'n bleken een betrouwbaar instrument (ICC 0,70 of hoger) bij toepassing door tenminste twee Test-VA'n.



### **Gedrag van de verzekeringsartsen met betrekking tot protocollen**

Het gedrag van de VA'n ten aanzien van protocollen in het algemeen en het protocol depressieve stoornis in het bijzonder komt aan bod in Hoofdstuk 4. Als basis voor het te bestuderen gedrag werd het ASE-model gebruikt. Voor het registreren van de persoonskenmerken van de deelnemende VA'n en het meten van de ASE determinanten van gedrag, en aanvullende factoren, werden vragenlijsten ontwikkeld. De meerderheid van de deelnemende VA'n gaf aan minimaal een aantal onderdelen van het protocol Depressieve stoornis te gebruiken. Het gedrag van de verzekeringsartsen ten aanzien van het protocol Depressieve stoornis werd bestudeerd door gebruik te maken van LISREL analyses. Het bleek dat VA'n zich in hun intentie het protocol Depressieve stoornis toe te passen lieten beïnvloeden door de mate van ervaren controle over het kunnen toepassen van het protocol, hun collega's, en de wijze waarop het protocol geïmplementeerd werd.

### **De evaluatie van de implementatiestrategie**

De resultaten van de implementatiestrategie worden gepresenteerd in Hoofdstuk 5. Het hoofddoel van deze studie was te evalueren in hoeverre de implementatiestrategie in staat was de 'guideline adherence' van de VA'n (de mate waarin de VA'n het protocol volgen) te verbeteren. Daarnaast werd de invloed van de implementatiestrategie op de kennis van de VA'n van het protocol Depressieve stoornis onderzocht. De ontwikkelde strategie werd onderzocht in een gecontroleerd experiment met een voor- en een nameting. In dit experiment werd de 'guideline adherence' van een groep VA'n die de implementatiestrategie kreeg, vergeleken met een controlegroep VA'n die een 'placebo training' kreeg. Alle deelnemende VA'n in het experiment beoordeelden de arbeidsongeschiktheid van vier verschillende cliënten met een depressie, twee voor en twee na de implementatiestrategie. Deze cliënten werden gespeeld door acteurs en werden gepresenteerd op video. De 'guideline adherence' in de VA-rapportages van de deelnemers werd bepaald door de Toets-VA'n met behulp van de PI'n (zie Hoofdstuk 3).

De VA'n die de implementatiestrategie hadden gekregen presteerden significant beter op de PI'n (gemiddeld 4,44 op een theoretisch gemiddelde PI-somscore van 1,00-5,67) dan de VA'n uit de controlegroep (gemiddeld 3,32). Een hogere PI-somscore van een VA-rapportage betekende dat de rapportage meer in overeenstemming was met het protocol Depressieve stoornis. De kennis van het protocol werd apart getest met een kennistoets. De VA'n uit de interventiegroep hadden na de implementatiestrategie een significant betere kennis van het protocol Depressieve stoornis dan de VA'n uit de controlegroep. De conclusie was dat bij de VA'n ten gevolge van de implementatie-strategie zowel de 'guideline adherence' als de kennis van het protocol Depressieve stoornis verbeterde.

### **Veranderingen in gedragsdeterminanten van verzekeringsartsen na de implementatiestrategie**

In Hoofdstuk 6 worden de veranderingen in gedragsdeterminanten van de VA'n met betrekking tot het protocol Depressieve stoornis ten gevolge van de implementatiestrategie beschreven. Met behulp van vragenlijsten die gebaseerd waren op het ASE-model, werden deze gedragsdeterminanten gemeten voorafgaande aan en drie maanden na het uitvoeren van de implementatiestrategie. De gedragsdeterminanten van de VA'n die de implementatiestrategie hadden gekregen, veranderde significant in positieve zin ten opzichte van de gedragsdeterminanten van de VA'n uit de controlegroep. De onderzochte determinanten attitude en intentie tot gebruik van het protocol verbeterden met 12%, zich toegerust voelen het protocol te gebruiken met 10%, en kennis en vaardigheden met 5%. Alleen de verandering in zelf gerapporteerde kennis en vaardigheden kon gerelateerd worden aan de geobserveerde 'guideline adherence' van de VA'n, zoals gemeten was met de PI'n. Deze gevonden relatie was echter zwak.

### **Aantal en ernst van de beperkingen en de mate van overeenstemming in arbeidsongeschiktheidsbeoordelingen tussen de VA'n bij toepassing van het verzekerings-genees-kundig protocol Depressieve stoornis**

In Hoofdstuk 7 hebben we onderzocht wat nu de invloed was van de implementatiestrategie op de beperkingen die de VA'n vaststelden in de Functionele Mogelijkheden Lijst (FML) bij cliënten met een depressie. De VA'n die deelnamen aan het experiment vulden voor alle vier de cliënten een FML in. Het bleek dat na de implementatiestrategie: 1) de VA'n meer beperkingen en zwaardere beperkingen gaven, en 2) de VA'n meer overeenstemden in hun arbeidsongeschiktheidsbeoordelingen dan de VA'n uit de controlegroep.

VA'n zouden zich moeten realiseren, dat zij met het volgen van een protocol zover meer informatie beschikken dan zij mogelijk gewoon zijn, hetgeen kan leiden tot het geven van meer en zwaardere beperkingen. En dat kan vervolgens een belemmerende werking hebben op de arbeidsparticipatie van de cliënt. Voor beleidsmakers is het van belang te weten dat het mogelijk is de uniformiteit in arbeidsongeschiktheidsbeoordelingen te vergroten door VA'n te trainen in het toepassen van een protocol.

### **De procesevaluatie van de implementatiestrategie**

Hoe de implementatiestrategie werd ontvangen door de deelnemende VA'n, wordt beschreven in Hoofdstuk 8. De 42 VA'n die deelnamen in onze studie waren in het algemeen zeer tevreden met de implementatiestrategie (een gemiddelde score van 7,7 op een 1-10 schaal). Zij waardeerden in het bijzonder de training en de bureaulegger. De meerderheid van de VA'n (81%) verwachtte direct na de implementatiestrategie een verbetering van hun arbeidsongeschiktheidsbeoordelingen bij cliënten met een depressie te kunnen realiseren, terwijl 86% van de VA'n een verbetering verwachtte van hun VA-rapportages. Na drie maanden beschouwde 96% van de VA'n de implementatiestrategie



nog steeds als nuttig. De tijd die nodig is om het protocol Depressieve stoornis toe te kunnen passen werd door sommige VA'n gezien als een obstakel voor de uitvoering in de praktijk. Tevens gaven sommige VA'n aan moeite te hebben met het veranderen van hun werkrouines voor het kunnen toepassen van het protocol Depressieve stoornis. Van de ongeveer 800 VA'n die bij het UWV werkten in 2008 deed slechts 5% mee aan deze studie.

### De algemene discussie

In Hoofdstuk 9 worden de antwoorden gegeven op de vragen die in de 'General Introduction' gesteld werden. De antwoorden op deze vragen zijn de hoofdbevindingen van deze thesis. Daarnaast worden in dit hoofdstuk methodologische beschouwingen en aanbevelingen voor verder onderzoek behandeld.

Het gebruik van de 'Intervention Mapping' methode voor de ontwikkeling, de planning en het vinden van de juiste benaderingswijze voor de implementatiestrategie is beschreven. Het ontwerp van de studie is besproken door de voor- en nadelen van de onderhavige effectstudie (efficacy design) te vergelijken met die van een effectiviteitsstudie (effectiveness design) tegen de achtergrond van de UWV context. Het originele plan de implementatiestrategie te evalueren in een effectiviteitsstudie op het UWV is verlaten vanwege praktische redenen. Daarom hebben we een 'efficacy' studie uitgevoerd in de vorm van een experiment in een gecontroleerde omgeving. De eerste uitkomstmaat, de mate waarin VA'n het protocol volgen ('guideline adherence'), met het bijbehorende meetinstrument, de performance indicatoren, is toegelicht. In een breder perspectief is stilgestaan bij de onderwijskundige aspecten van de ontwikkelde implementatiestrategie. Het gedrag van de VA'n ten aanzien van protocollen is bediscussieerd aan de hand van psychologische gedragsmodellen en literatuur. Aansluitend is onze implementatiestrategie vergeleken met wat in de literatuur wordt beschreven als effectieve implementaties. Obstakels in de implementatie van protocollen zijn in het algemeen een belangrijk onderwerp in studies, en daarom hebben we een paragraaf gewijd aan de mogelijke obstakels in onze implementatiestrategie.

### Conclusies en implicaties

Een implementatiestrategie voor het verzekeringsgeneeskundig protocol depressie werd ontwikkeld en geëvalueerd. De werkzaamheid van deze implementatiestrategie werd aangetoond. VA'n die de implementatiestrategie kregen bleken het protocol Depressieve stoornis beter te volgen, en kenden het protocol beter dan VA'n die alleen de gebruikelijke implementatie bij het UWV hadden gekregen. De VA'n waardeerden de implementatiestrategie, zoals bleek uit de procesevaluatie en uit een positieve gedragsverandering ten aanzien van het protocol Depressieve stoornis. Ten slotte verbeterde de overeenstemming tussen de VA'n bij de beoordeling van videocasus van een cliënt met depressie, hetgeen wijst op een toename van uniformiteit in de arbeidsongeschiktheidsbeoordelingen van cliënten met een depressie.

De algemene conclusie is dat we erin zijn geslaagd een rijk geschakeerde implementatiestrategie te ontwikkelen voor het protocol Depressieve stoornis. Echter, deze implementatiestrategie dient nog geëvalueerd te worden in de praktijk. Daarnaast werd de implementatiestrategie in dit onderzoeksproject slechts geëvalueerd op het niveau van de VA, bovendien in een gecontroleerde omgeving. Belangrijke niveaus voor een implementatie, zoals het niveau van de organisatie of dat van de cliënt, werden niet meegenomen in dit onderzoeksproject.

De resultaten van dit proefschrift hebben praktische implicaties voor VA'n, voor beleidsmakers en voor opleidingsprogramma's gericht op VA'n. Voor VA'n: deze implementatiestrategie kan de kwaliteit van de VA-rapportage verbeteren, hetgeen het werk van de VA transparanter maakt voor anderen; daarnaast kan de uniformiteit in de arbeidsongeschiktheidsbeoordelingen toenemen. Voor de beleidsmakers van het UWV: deze implementatiestrategie zou landelijk ingevoerd moeten worden. Daarnaast kan deze strategie ook worden aangepast voor andere verzekeringsgeneeskundige protocollen. Het monitoren van VA-prestaties op het volgen van protocollen is nu mogelijk geworden. Voor opleidingsprogramma's gericht op VA'n: de vertaalslag van 'evidence-based medicine' uit de protocollen naar de dagelijkse praktijk van de VA kan worden bereikt door gebruik te maken van ervaren VA-trainers en met hulp van realistische casuïstiek uit de praktijk van het UWV. Deze implementatiestrategie voor het protocol Depressieve stoornis draagt bij aan de kwaliteitsslag die de verzekeringsgeneeskunde maakt.

# **Dankwoord**



## Dankwoord

Toen ik in de zomer van 2007 begon als junior onderzoeker bij het KCVG werd ik gevraagd mezelf te introduceren door een stukje in het personeelsblaadje van het EMGO Instituut te schrijven. Dat was een helder stukje, waarin ik mijn verwachtingen voor de komende jaren uitsprak. Helaas kan ik dat stukje niet meer vinden, maar de strekking herinner ik mij nog wel. Ik was toen heel blij met de term "junior onderzoeker", aangezien ik op dat moment de 50 jaar passeerde, en ik slechts begon te worden na 20 vermoeiende jaren in de uitvoering van de sociale wetgeving. Hier ging een nieuwe wereld voor mij open, en ik voelde mij als Kuifje in onderzoeksland. Het zou een spannend avontuur worden met vele uitdagingen.

Om te beginnen het onderwerp "Protocol Depressie". Probeer daar maar eens wat leuks van te maken. Tijdens het sollicitatie gesprek werd mij al gevraagd "**Hoe denkt u die artsen te leren dat protocol toe te passen**". Ik dacht alleen maar: die verzekeringsartsen hebben momenteel hun buik vol van al die protocollen. Dus ik zei: "Dat zie ik wel, maar ik denk wel dat ik het kan". Nu was ik de tweede promovendus op dit onderzoeksproject. Toen ik begon was het mijn ding nog niet. Het werd mijn project op het moment dat het logo Protocol Depressie verscheen. En dat logo gebruikte ik in de vele presentaties die ik hield. Door anderen te overtuigen. Zo begon ik zelf te geloven in de smiley die in de P van het protocol stond. Langzaam begon ik enige grip op het onderwerp te krijgen. Onder het adagium van mijn ouders "jij redt je wel" ging ik aan de slag en zo geschiedde. Het leuke van zo'n project is dat je verschillende rollen kunt aannemen: kamergeleerde of nerd, projectmanager, administratief medewerker, collega, presentator, dokter, leraar, en leerling of gezel. Tijdens het maken van de video's kon ik ook nog in de rol van scriptwriter en regisseur kruipen. Graag had ik nog een cliënt met een depressie gespeeld maar dat was niet objectief genoeg voor het onderzoek. In al die rollen heb ik intensief met velen samengewerkt met als uiteindelijk resultaat dit proefschrift. Daarvoor wil ik al die mensen met wie ik heb samengewerkt bedanken. Hier volgen ze:

### Promotoren en copromotor

Ik ben dankbaar dat ik in deze periode deel heb uit mogen maken van een goed en inspirerend team van deze zeer verschillende mensen met een passie voor de wetenschap.

Ton, de eerste copromotor, die altijd een paar stappen vooruit is en dat koppelt aan een fantastische inzet en drive. Waarvoor ik veel respect heb. In het begin kon ik hem amper volgen en had ik moeite hem bij te benen. Het begrip moest bij mij nog groeien en Ton hield dat goed in de gaten. Nu nog steeds weet hij eerder wanneer ik iets wel of niet snap dan ikzelf, hetgeen opmerkelijk is. Ook al " proefden we elkaar wel eens de

nieren", in Ton's woorden, hebben we zeer vruchtbaar samengewerkt. Ik heb altijd op hem kunnen rekenen. En rekenen dat kan hij.

Ton: "*Feico, dit kun je niet intuïtief oplossen, dit moet je gewoon uitrekenen. Kijk hier, ik teken een modelletje, het is echt heel simpel*".

Han, eerst de tweede copromotor en later promotor, is naast wetenschapper ook arts en het was heel goed om een arts in het team te hebben. Het onderzoek ging ten slotte over artsen, bij wie we moesten aansluiten om het van de grond te krijgen. Han wist met zijn ruime wetenschappelijke ervaring, zijn pragmatische inslag, en niet te vergeten zijn feilloos gevoel voor verhoudingen, precies op het goede moment bepaalde zaken snel voor elkaar te krijgen. Hetgeen ik een kunst vind in de wereld van de wetenschap.

Han: "*Feico, ik mis hier wat kopjes. Als je het nu zo en zo indeelt met die kopjes, en dat één op één laat terugkomen in de Methode en de Results en even refereert aan die en die dan krijg je het veel eerder gepubliceerd*".

Allard, de promotor, degene die het laatste woord heeft en dat ook waar maakt. Allard was als promotor goed aanwezig binnen het team, heel benaderbaar en betrokken. Daarnaast sportief en relativerend, maar scherp op het resultaat wanneer het nodig is. Een aanvoerder zoals je het zou wensen als teamspeler.

Allard: "*Feico, hou nou eens op met creatief doen, een wetenschappelijke tekst is gortdroog, en niet te vergeten eenduidig. Dit kan je wel vinden maar dat kan je echt niet zo opschrijven. Dat is jouw mening, dat is geen wetenschap*".

Het is duidelijk dat ik vooral veel moest af- en daarna aanleren. Heren bedankt hiervoor.

### De leescommissie

Ik wil de leden van de leescommissie bedanken voor hun aandacht en tijd die zij hebben besteed aan het beoordelen en goedkeuren van mijn proefschrift: prof.dr. A.T.F. Beekman, prof. dr. J.J.L. van der Klink, prof. dr. C.T.J. Hulshof, prof. dr. M.J.P. Wensink, dr P. Eken en dr. P.B.A. Smits.

### De deelnemers

Zonder deelnemers geen onderzoek. De deelnemers zijn verzekeringsartsen die werken voor het UWV. Het zijn mijn collega's. Zij hebben tijd vrij gemaakt om iets te willen leren. Daarvoor kwamen ze uit het hele land en zelfs uit België naar Amsterdam. Zij hebben hard gewerkt met een voor mij opvallende inzet en toewijding. De resultaten van dit onderzoek zijn in feite hun prestaties. Enkele citaten van de deelnemers.



*“Waarom krijgen we niet zo’n bureaulegger voor alle protocollen?”*

*“De verschillen tussen de verzekeringsartsen zijn op dit punt (beoordelen van cliënten met een depressie) enorm. Het is goed dat daar op getraind wordt”. “Dit vond ik een hele leuke dag”.*

### **De verzekeringsartsen op de DVD’s**

Theya Njoo en Hans Goossens.

Zij hebben zich op bijzondere wijze ingezet voor dit project, niet alleen door zich te laten filmen in de uitoefening van hun werk, maar door ook op mijn instructies in de geest van het protocol te werken. Bovendien kregen ze niet zo maar wat cliënten voorgeschoteld, maar pittige casuïstiek. In die casuïstiek had ik me als scriptwriter juist lekker uitgeleefd.

Verzekeringsarts: *“Drinkt u?”* Cliënt: *“diepe zucht, Nou ja, een paar biertjes misschien, om nou te zeggen dat ik drink?”*

Verzekeringsarts: *“Bent u nu al bij het re-integratiebureau geweest?”* Cliënt: *“Tja, ik zou daar nog een keer heen gaan om koffie te drinken of zo, maar daar is het nooit meer van gekomen.”* Verzekeringsarts: *slaakt een diepe zucht*

Verzekeringsarts: *“Kunt u genieten van het leven?”* Cliënt: *geen reactie, staart naar de grond.* Verzekeringsarts: *“Hmmm”.*

### **De acteurs en actrices op de DVD’s**

Ilke Turpijn, Ellenor Spreeuw, Vanessa de Boer, Dirk van der Pol en Jorick Jochims. Zij acteerden als cliënten in dit onderzoek en zij wisten dat op zeer realistische wijze te doen. Dit proefschrift gaat over een protocol dat wordt toegepast op mensen. Daarom staat actrice Ilke Turpijn, die een vrouw met depressieve klachten speelt, op de omslag van dit proefschrift met de banner van het Protocol depressie over haar heen.

### **De Test-Verzekeringsartsen**

Paula Eken, Theya Njoo, Monique Stroomer, Eric van der Jagt, Ron de Vink, Hans Goossens, Jan van Oort en Dirk van Latenstein.

Zij hebben zich extra ingezet voor dit onderzoeksproject. Deze verzekeringsartsen hebben na eerst twee dagen geoefend te hebben, wekenlang de rapportages van de deelnemers getest op het toepassen van het protocol. Deze klus moest met de nodige toewijding en nauwgezetheid geklaard worden. Hier enkele uitspraken uit de test sessies van deze zeer betrokken groep verzekeringsartsen:

*“Sommige rapportages zijn echt prut”. “Bij andere rapportages zie je dat de structuur ergens wel klopt maar is het inhoudelijk toch weer prut”. “Ik ben al heel blij dat een verzekeringsarts het noemt in de rapportage”. “Ja, OK het staat er wel, maar is het nu beargumenteerd? Nee dus”. “De meeste rapportages zijn heel behoorlijk”.*

### **Psychiater Mieke Hassing**

Zij wist de verzekeringsartsen te boeien met een prikkelende presentatie over een patiënt met een depressie uit haar eigen praktijk, die ook cliënt bij het UWV was. Een mooi voorbeeld van samenwerking tussen de specialist en het UWV.

### **De onderzoeksassistentes**

Sietske Tamminga, Karlijn van Beurden en Karin Groenewoud.

Voor mij zijn ze volstrekt onmisbaar. Vooral op het gebied van redactionele vaardigheden. Dames bedankt.

*“Feico, wat heb je nu weer met dat document uitgehaald? Overal zie ik spaties die er niet horen en bij de referenties is het helemaal een zootje”. “Ik probeerde het juist netjes te krijgen, maar het werd alleen maar erger”.*

### **De KCVG collega’s**

De KCVG collega’s hebben mij de afgelopen jaren enorm gestimuleerd. Ik vond het leuk met hen te sparren over mijn onderzoek en dat van hen. Het leuke aan deze contacten is dat je er niet alleen professioneel wat aan hebt, maar ook als mens. Die momenten dat je het even zwaar hebt, dan is er altijd wel een collega van het KCVG in de buurt bij wie je een willig oor vindt. Speciaal wil ik noemen mijn paranimfen Jolanda van Rijssen en Rob Kok. En natuurlijk Sylvia Vermeulen en Diederike Holtkamp.

### **De EMGO+ collega’s**

In ieder geval kamergenoten, Sonja Schut, Sjaak Broersen, Peter van Muijen, Eva Bouwsma, en verder iedereen die ik er ken. Hoewel ik het gebouw nogal deprimerend vind, werd dat ruimschoots gecompenseerd door de levendige en enthousiaste collega’s bij wie ik altijd terecht kon, voor een vraag of nog belangrijker voor het lenen van de koffiekaart, die ik al na een week kwijt was.

### **Jacques Koeweiden ([www.koeweidenpostma.com](http://www.koeweidenpostma.com))**

Jacques ontwierp het logo voor dit onderzoeksproject.

Jacques zei: *“Vertel eens wat over dat onderzoek van je”.*

Ik kletste wat over het protocol depressie. Hij krabbelde ondertussen wat op papier en binnen een minuut had hij het logo klaar. Perfect. Om jaloers op te zijn. Zo’n onderzoeksproject daarentegen is een proces van jarenlang wikken en wegen, rekenen, plannen, veel schrijven en vervolgens weer comprimeren tot de essentie. En hij boekt resultaat binnen een minuut. Jacques, bedankt.



### **De NSPOH**

In de persoon van Tineke Woldberg en Brigitte Fennis. Tineke zag direct het belang van mijn onderzoek in, hetgeen resulteerde in een vlotte en prettige samenwerking. Samenwerken brengt je verder. Brigitte verzorgde de broodnodige ondersteuning.

### **UWV**

Ik ben trots en verheugd dat het UWV dit project mogelijk heeft gemaakt. Na meer dan twintig jaar in de productie gewerkt te hebben kreeg ik de kans mee te werken aan deze kwaliteitsslag. Daarbij wil ik de cliënten van het UWV niet vergeten. Eigenlijk vind ik dat de cliënt in dit project te weinig aan bod is gekomen, maar het gaat er indirect wel degelijk om.

Verzekeringsartsen oordelen over cliënten en die zijn er bij gebaat dat de arts dat zorgvuldig doet. De casuïstiek die gebruikt is in dit onderzoeksproject is gebaseerd op cliënten uit de praktijk van het UWV. Speciaal wil ik de afdeling Opleidingen van het UWV bedanken, die hebben bijgedragen aan de financiering van dit proefschrift.

*Na afloop van de Dam tot Damloop (onderdeel bedrijvenloop) die ik liep voor het UWV, zat ik in de bus terug naar Amsterdam naast een anonieme medeloper. Hij vroeg: "En voor welk bedrijf loopt U?". "Voor het UWV", was mijn antwoord. Hij was even stil, keek me aan en zei: "Ik hoop niet dat ik ooit met u te maken krijg". Toen was ik even stil. "Dat zeg ik ook tegen mensen van het Rode Kruis, hoor", voegde hij toe. "Goed dat ze er zijn, maar je moet er niet mee te maken krijgen".*

### **Familie en vrienden**

Het is gebruikelijk dat nu de tekst volgt, dat ik het zonder de steun van intens meelevende vrienden en familie niet had gered. Nu, het ligt voor mij toch wat anders. Laat ik zeggen dat ik hoop dat ik mijn vrienden en familie niet te veel belast heb met dit toch wel taaie onderwerp. Eigenlijk had ik het schrijfwerk van dit project willen voltooien als een monnik in een klooster. Volledige concentratie en toewijding op het proefschrift en verder niets. Ik vond het soms moeilijk het werken aan dit project te combineren met mijn gewone leven. Het liet me gewoon niet los. Ik wil speciaal mijn meest dierbaren, mijn vrouw Giannina en mijn dochter Giulia, bedanken die dit goed aanvoelden.

# **About the author**

**Feico Zwerver**



## About the author

Feico Zwerver was born on September 25th 1957 in Alkmaar, the Netherlands. After completing secondary school at the Kennemer Lyceum in Overveen in 1977, he studied history of art at the University of Amsterdam in 1978. After fulfilling his military obligations from 1978 till 1979 he studied Medicine at the University of Amsterdam. He received his medical degree in 1987. He then specialized in social insurance medicine and he was registered in 1994.

In 1988 Feico started working as an insurance physician for the GMD and later for the UWV (Dutch National Institute for Employee Benefits Schemes). In 2007 he was detached as a junior researcher for 80% of his time to the Dutch Research Center for Social Insurance Medicine, a collaboration between the VU University Medical Center, the Academic Medical Center, The University Medical Center Groningen and UWV. For his PhD project, he worked at the Department of Public and Occupational Health, EMGO Institute for Health and Care research, VU University Medical Center, Amsterdam. The post-graduate course for insurance physicians, which embedded the experiment of this thesis, was realized in cooperation with the Netherlands School of Public & Occupational Health (NSPOH). Feico takes an active part in teaching medical students at the VU University Medical Center, and teaching insurance physicians at both the NSPOH and the UWV. Currently, he is working as an insurance physician in practice and as an insurance physician trainer at the UWV.

Feico is married to Giannina, and they have one daughter Giulia (9). Hobbies are running, tennis, ski, fly-fishing and chess.

