

INVESTING IN EMPLOYABILITY INTERVENTIONS ?

methodological challenges
and economic evaluation results

CINDY NOBEN



The studies presented in this thesis were performed at the School for Public Health and Primary Care (CPAHRI), Department of Health Services Research (HSR) at Maastricht University.

The research presented in this thesis was funded by the Dutch Foundation Institute Gak as part of the program 'Arbeid & Zorg'.

Printing of this thesis financially supported by CAPHRI HTA program and ACT désiron is gratefully acknowledged.

Copyright © Cindy Yvonne Gerarda Noben, Maastricht, 2015

ISBN 978-94-90411-07-7

Cover design by Els Maurissen

Printed by Datawyse / Universitaire Pers Maastricht

All rights reserved. No part of this thesis may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage or retrieval system, without prior written permission of the holder of the copyright.

Investing in employability interventions?

Methodological challenges and economic evaluation results

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,
op gezag van de Rector Magnificus, Prof. dr. L.L.G. Soete,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen,
op maandag 6 juli 2015 om 12.00 uur

door

Cindy Yvonne Gerarda Noben
Geboren op 1 juni 1986 te Tongeren

Promotores

Prof. dr. S.M.M.A Evers

Prof. dr. F.J. Nijhuis

Copromotor

Dr. A.E. de Rijk

Beoordelingscommissie

Prof. dr. D. Ruwaard (voorzitter)

Prof. dr. W.B.F. Brouwer (Erasmus University Rotterdam)

Prof. dr. C.S. Dewa (University of Toronto)

Prof. dr. A. de Grip

Prof. dr. IJ. Kant

*Aan mama,
omdat ze altijd in mij is blijven geloven*

TABLE OF CONTENTS

Chapter 1	General introduction and outline	9
PART I Methodological challenges		
Chapter 2	Quality appraisal of generic self-reported instruments measuring health-related productivity changes: A systematic review.	21
Chapter 3	Patient self-reported and administrative data used for cost estimations of healthcare resource use: A systematic review of methodological reporting quality and validation evidence.	71
Chapter 4	Design of a trial-based economic evaluation on the cost effectiveness of employability interventions among work disabled employees or employees at risk of work disability: The CASE-study.	103
PART II Existing practices		
Chapter 5	Comparative cost effectiveness of two interventions to promote work functioning by targeting mental health complaints among nurses: Pragmatic cluster randomized trial.	125
Chapter 6	Protecting and promoting mental health of nurses in the hospital setting: Is it cost effective from an employer's perspective?	147
Chapter 7	A pilot economic evaluation to improve a web-based employability intervention for employees.	161
PART III Innovative interventions		
Chapter 8	Economic evaluation of a new organizational RTW intervention to improve Cooperation between Sick-listed employees and their Supervisors (COSS): A field study.	185
Chapter 9	Economic evaluation of an intervention program with the aim to improve work productivity for workers with rheumatoid arthritis: A six months interim analysis.	205
Chapter 10	Short-term cost effectiveness and cost utility of a training program to increase GP awareness to work-related problems: A randomized controlled trial.	227

PART IV Discussion

Chapter 11	General discussion	251
Chapter 12	Valorization of research results	271
	Summary	281
	Samenvatting	287
	Dankwoord	291
	Curriculum Vitae	301
	Scientific contributions, grants & awards	303

CHAPTER

1

General introduction and outline

Introduction

Despite broad societal consensus on the growing importance of sustainable employability, health problems continue to limit labor participation. The health care sector seems rather unresponsive to work and health-related issues, and employers and employees remain unable to fulfill their mutual responsibilities to create healthy work environments [1, 2]. These key parties have particular difficulties addressing the structural problems in this area [3, 4]. The mismatch between health-related activities and the deployment of personnel in the workplace contributes to absenteeism rates and loss of productivity [2, 5]. There is strong potential for work-related health interventions to improve labor participation [6].

Like other Western countries, the Netherlands has been struggling with structural barriers to reduce the negative effects of health problems on work and the work context [7, 8]. The negative impact of health problems in the work setting can hamper labor participation and potentially lead to lost productivity, early retirement and work incapacity, and higher social security costs [9, 10]. In addition, more recent developments, such as the rising pension age, call for innovative methods to increase sustainable employment of employees with health problems. Despite various research projects and significant changes in the legal framework for occupational health and social security¹ the relationship between health and labor is far from optimal. Initiatives and experiments remain too vague and fragmented and employers are not inclined to pay for them if they are not proven effective. More insight into the effectiveness and cost effectiveness of the initiatives targeting this important problem is needed. Cost effective interventions are attractive for the individual employee, the employer, the broader occupational health practice, but also society as a whole because illness duration, care consumption, and the costs of lost productivity will decrease when such interventions are utilized. Furthermore, cost effective employability interventions are expected to increase labor participation and public health [11].

In light of this, a more structural approach is necessary to develop appropriate responses when health problems impede or threaten employability and labor par-

¹ The Dutch social security system promotes active participation and work ability via:

(i) The Gatekeeper Improvement Act, which was introduced in 2002 to prevent (long term) sickness absence and reduce the inflow of beneficiaries in disability pension schemes. The new law stated a mutual responsibility for both the employer and employee to improve return to work efforts. In the first two years, the employer is provided with salary and mandatory reintegration obligations, and the employee has to actively participate in the mandatory reintegration obligations and should accept the work activities the employer provides. Only if no reintegration is achieved after these two years, the employee can apply for long term disability benefits.

(ii) The Work Capacity Act, which replaced the Work Disability Act in 2005. The focus shifted from 'what the employees should be compensated for' to 'what the employee is able to do'. This new approach of health-related work functioning, whereby employees are supported to enrich their abilities to meet the work demands given their health state, no longer focused on return to work but on staying at work.

ticipation. Important first steps have already been taken. The awareness of employers, employees, and health care providers about the relationship between work and health has increased due to recent social developments, such as the aging population. Furthermore, aside from work-related or work-caused health problems, sustainable employability has also gained more prominence in connection with health issues arising outside of the work context. Several attempts are being made to develop and implement interventions that focus more on the work context, and thereby on labor participation, regardless of the cause of the health complaints (work or non-work related). However, most interventions remain focused on the health problems caused by the work, resulting in an insufficiently dynamic approach [10, 12-14]. To illustrate this, shifts in the nature of work – from heavy physical labor towards knowledge-based and people-oriented work – can cause, but also change, work-related disorders. More attention needs to be paid to mental disorders impeding sustainable employability (e.g. burnout) and the strong separation of mental health care and occupational health care [15]. Additionally, growing awareness also emerged about the need to better equip employees and their employers with specific skills. The early detection of health problems that potentially hamper work, and the prevention of reduced labor participation, are needed to appropriately respond to health problems impeding employability and labor participation [16].

This PhD-study is part of a broader research program, called ‘The Work and Care research program’². The Work and Care research program was set up to assess the complete path from symptoms and calling in sick, to working on medical and functional recovery, early return to work, or – if that failed – reintegration within two years. The research program focused on how employers and employees can gain greater control in this process and on ways to explicitly involve professionals in the health sector to achieve employability. It was stated that today’s society must offer people opportunities (i) to maintain their work, with or without adjustments, (ii) to return to suitable work, (iii) to make a start with work that suits their physical and mental capabilities.

In that context, this PhD-thesis provides insight into both methodological and empirical (experimental) analyses of interventions aimed at achieving sustainable employability for employees with health problems that potentially impede their labor participation. This thesis mainly focuses on the underlying health economic aspects, and thus on the cost effectiveness of interventions aimed to achieve sustainable employability. The assessed interventions target employees who developed complaints and/or a disease impeding their labor participation, and for that matter also their potential to achieve sustainable employability. This might involve either acute injuries or a chronic illness or disability. The objectives of the studied

² The Work and Care national research program is commissioned and financed by the Foundation Institute Gak (In Dutch: Stichting Instituut Gak)

interventions include prolonging successful and sustainable employability for employees with health problems (for example by preventing absenteeism among employees with a chronic disease) and increasing compliance with the existing legislation when performing labor (for example by increasing cooperation among employees and employers).

Sustainable employability implies that employees throughout their working life continuously have real opportunities, a set of capabilities, and the necessary conditions that allow them to achieve valuable work functioning in current and future work with preservation of health and welfare. This implies a work context that facilitates them, as well as the attitude and motivation to exploit these opportunities [17].

Employees are a heterogeneous group, involving people of all ages, both men and women, with different levels of education and work experience, and varying work barriers spanning the whole spectrum of psychiatric, psychological and physical limitations. Accordingly, it is important to study a diverse range of employability interventions. The included employability interventions in this PhD-thesis therefore allow for the economic evaluation of any employability intervention, regardless of the population they were developed for. To illustrate this, employability interventions provided in the primary and secondary health sectors, as well as those delivered within the workplace, are evaluated. The analyses of these interventions lead to an understanding of the quantitative effects (including reduced absenteeism, health care utilization, and improved work abilities) for employees to continue to participate in labor, or to increase their sustainable employability.

Throughout this PhD-thesis, two different existing employability interventions and three innovative employability interventions are studied. Each employability intervention included in this PhD-thesis – new or existing – has a common goal: to increase the potential of achieving sustainable employability among employees with health complaints.

Employability interventions aim to improve or restore the sustainable employability of employees with health problems or those at risk of developing health problems that might potentially hamper the employee's work

As described above, demographic changes and political decisions influence the achievement of sustainable employability. Furthermore, companies need to remain productive and competitive in the global market and healthy work forces play an important role in achieving these outcomes. However, ensuring workforce health is not a simple task due to the diverse range of stakeholders involved at the individu-

al, workplace, and/or societal level [4] and the local and national arrangements regarding disability benefits and compensation schemes for illness and disease³. This complexity is especially relevant for countries where not only the employees, but also other parties and systems have responsibilities for and contribute to the work disability context, as is the case in the Netherlands [18]. Because of the increasing responsibilities (and thus also potential risks) that employers and employees face in the Dutch 'participation society', this research is of social value.

To properly assess the added value of employability interventions aimed at achieving sustainable employability for employees with health problems, evidence comprising economic evaluations and cost effectiveness results are needed. This PhD-thesis generates new evidence on the cost effectiveness of the assessed employability interventions.

An economic evaluation can be defined as the comparative analysis of alternative courses of action in terms of both their costs and consequences [19].

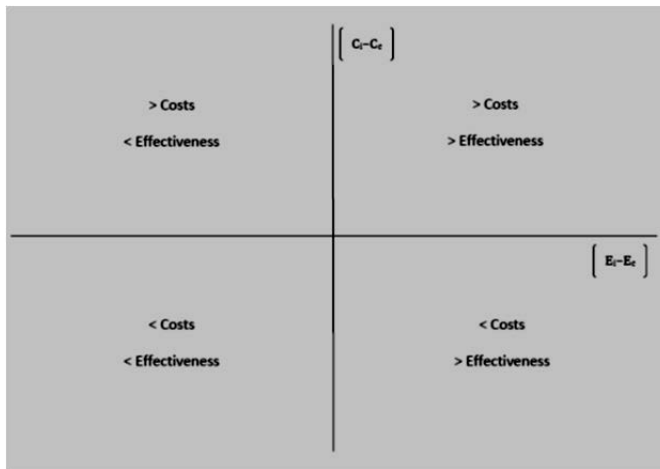
When conducting an economic evaluation it is necessary to identify the relevant costs and effects attributable to the intervention in question in a comparative analysis. Claiming that an intervention is cost effective is equivalent to saying that the extra effect is worth the extra cost [19]. Identifying the costs and effects is often influenced by the chosen perspective. Because employability issues affect multiple stakeholders, such as employees and their families, but also firms, insurers, healthcare providers, and even society as a whole, considering a broad societal perspective is essential [20]. Choosing the appropriate kind of evaluation to assess employability interventions can be challenging because of the range of stakeholders involved. Therefore, different types of economic evaluations exist, whereby the key consideration is the trade-off that matters most to the decision maker.

The three main types of economic evaluations used throughout this PhD-research are cost benefit-, cost effectiveness-, and cost utility analyses. All three analyses express costs in monetary currencies, but differ in effect countenance. An important feature within economic evaluations is measuring and valuing the costs and effects. When all costs and effects relevant for the comparative analysis are identified, measured and valued, they can be averaged across all participants in both groups (often illustrated as the intervention _(i) and control _(c) group). The mean costs (C) and mean effects (E) are obtained for each group in order to calculate the

³ Disability pension schemes provide disabled employees with income replacement benefits after sickness absence benefit ends. They are also called disability benefits, and the level and eligibility vary between and within countries. Eligibility criteria to receive disability benefits and the length of coverage differ from one disability insurance plan to another. In the Netherlands, the social insurance physician (if there is still work-ability), and a vocational expert (in Dutch: arbeidsdeskundige) decide on the percentage of loss of income due to work disability for which the employee will be partly compensated for.

incremental cost effectiveness ratio (ICER). The ICER represents the difference in costs divided by the difference in effects: $ICER = (C_i - C_c) / (E_i - E_c)$. To deal with sampling uncertainty, bootstrapped ICERs can be calculated and plotted on a cost effectiveness plane (see Figure 1) whereby the X-axis shows the difference in effectiveness and the Y-axis the difference in costs for both groups. When the majority of the ICERs are located in the south east quadrant of the cost effectiveness plane, indicating fewer costs and more effects, the intervention should be adopted. To deal with uncertainty related to the assumptions made, or to test the robustness of the findings, sensitivity analyses can be conducted [21-23].

Figure 1. Cost effectiveness plane



An economic evaluation assessment for interventions aimed at obtaining sustainable employability for employees with health problems requires a different approach than an economic evaluation of health care programs [20, 24]. To illustrate this, the policy arena for social security and labor legislation for occupational health are both complex, each having a diversity of involved stakeholders [24, 25]. These stakeholders potentially have different perspectives on which costs and effects should be quantified and expressed in monetary values. Importantly, the burden of the costs and the benefits of the consequences may be borne by different stakeholders in the system [24]. Furthermore, the benefits in this field are not primarily focused on health, but instead focus on such outcomes as productivity, work ability, remaining at work, preventing presenteeism and absenteeism. Consequently, no generic methodology accurately captures the full costs and consequences of interventions in the field of occupational health [26]. Therefore, it is necessary to develop an economic evaluation methodology expressly for use in the evaluation of sustainable employability interventions, accounting for the diverse stakeholders' responsibilities.

This PhD-thesis aims to fill this gap by providing insight in how the methodology underlying economic evaluations within occupational health can be improved. It does this by focussing on the psychometric issues at stake when measuring necessary aspects (such as productivity changes or health care utilization) when conducting an economic evaluation in the field. Furthermore, this PhD-thesis demonstrates how an economic evaluation can contribute to the development and implementation of employability interventions.

To summarize, this PhD-thesis thus aims to provide new insight into:

- the methodological challenges when conducting economic evaluations of employability interventions,
- the application of this methodological knowledge to different employability interventions and the thereby generated evidence of the economic evaluation (cost effectiveness) of different existing and innovative interventions aimed at achieving sustainable employability for employees with health problems.

By combining the exploration of methodological challenges and the economic evaluation results, an answer is provided to whether or not it is recommended to *'invest in employability interventions.'*

Outline of the PhD-thesis

Following the introduction, this PhD-thesis is subdivided in four main part: Part I represents the *methodological challenges* when conducting economic evaluations of employability interventions, Part II applies this methodological knowledge to several *existing practices* aimed at achieving sustainable employability for employees with health problems, Part III also applies this methodological knowledge, and generates evidence of the cost effectiveness of *innovative interventions* aimed at improving or restoring sustainable employability for employees with health problems, Part IV presents a *discussion* of the main results and issues, and provides recommendations and identifies challenges associated with implementing this evidence.

In **Part I, Chapters 2 to 4**, measurements for health-related productivity and resource use are critically assessed to improve the methodology of economic evaluation and a methodological design for the economic evaluation of diverse employability interventions is presented. In recent decades, a debate has developed about how productivity costs should be monetarily valued (e.g. friction cost-, or human capital approach). Before proceeding further with this discussion, correctly measuring health-related productivity due to health problems is essential. Until now, methods of lost productivity and associated costs have varied substantially. **Chapter 2** presents a critical quality appraisal of the measurement properties of instruments measuring lost productivity. This appraisal is needed to calculate health-related productivity costs in a valid and reliable manner. Additionally, when calculating costs for health economic evaluations, insight into healthcare resource use is also crucial. Because healthcare resource quantities can be defined in different ways, a robust data collection method is critical. **Chapter 3** provides insight into

the validation evidence and the methodological reporting quality of self-reported and administrative healthcare resource use measurement methods. These insights are warranted before estimating health care resource utilization costs. **Chapter 4** describes the study design of the CASE-study, which refers to 'Cost Effectiveness Sustainable Employability'.

In **Part II, Chapters 5 to 7**, several existing practices are critically assessed to generate evidence on their abilities to achieve sustainable employability for employees with health problems by means of economic evaluations. The existing employability interventions are those already operational in the field of occupational healthcare during, or at the start of the PhD trajectory. **Chapter 5** presents the comparison of two strategies to promote work functioning among nurses at risk of mental health complaints and **Chapter 6** provides an in-depth analysis to detect whether this preventive intervention in the work setting might also be cost saving from an employer's perspective. In **Chapter 7** the results of a pilot economic evaluation for a web-based employability intervention are used to guide the improvement of the intervention content and study design.

In **part III, Chapters 8 to 10**, employability interventions which use an innovative, systematic approach in order to support work participation in employees with health complaints are economically evaluated. The innovative employability interventions have been newly developed for this research and not yet implemented in the occupational health care practice. **Chapter 8** reports on the cost effectiveness, -utility, and -benefit of an innovative intervention called COSS (COoperation between Sick-listed employees and their Supervisors) to improve cooperation between employees with health problems and their supervisor. **Chapter 9** presents the results of a short-term economic evaluation of an intervention program aimed at improving work productivity for employees with rheumatoid arthritis (RA). In **Chapter 10**, a short-term economic evaluation provides insight in the abilities of general practitioners (GPs) to address work-related problems among their patients.

In **Part IV, Chapter 11 to 12**, a general conclusion and valorization of the findings are provided. Practical and methodological issues in occupational health research, and implications and recommendations for practice and future research are presented.

References

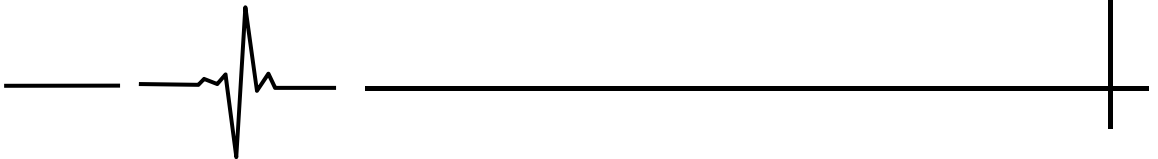
1. OECD. Transforming Disability into Ability. Policies to promote work and income security for disabled people. Paris: OECD Publications, 2003,2006,2008,2010.
2. Kremer AM, Steenbeek R. Avoidable sickness absence in a Dutch working population. *J Occup Rehabil*. 2010 Mar;20(1):81-9.
3. Sickness, Disability and Work: breaking the barriers 2010 [cited 2014]. Available from: http://ec.europa.eu/health/mental_health/eu_compass/reports_studies/disability_synthesis_2010_en.pdf.
4. Friesen MN, Yassi A, Cooper J. Return-to-work: The importance of human interactions and organizational structures. *Work*. 2001;17(1):11-22.
5. Hesselting JH, Houtman I, vd. Bossche S. Trends in ziekteverzuim (Trends in sickness absence). The Hague, The Netherlands: Sociaal Cultureel Planbureau; 2012.
6. Waddell G, Burton AK. Is work good for your health and well-being? In: TSO, editor. London2006.
7. Hoefsmit N, de Rijk A, Houkes I. Work resumption at the price of distrust: a qualitative study on return to work legislation in the Netherlands. *BMC Public Health*. 2013;13:153.
8. OECD. Sickness, Disability and Work: Breaking the Barriers: A Synthesis of Findings across OECD Countries. 2010.
9. Prins R. Sickness absence and disability: An international perspective. In: Loisel P, Anema JR, editors. *Handbook of Work Disability: Prevention and Management*. New York: Springer Science+Business Media; 2013.
10. Meerding WJ, Koopmanschap MA, Severens JL, Burdorf A. Health problems lead to considerable productivity loss at work among workers with high physical load jobs. *J Clin Epidemiol*. 2005 May;58(5):517-23.
11. Noben CY, Nijhuis FJ, de Rijk AE, Evers SM. Design of a trial-based economic evaluation on the cost effectiveness of employability interventions among work disabled employees or employees at risk of work disability: The CASE-study. *BMC Public Health*. 2012 Jan 18;12(1):43.
12. Hoeijenbos M, Bekkering T, Lamers L, Hendriks E, van Tulder M, Koopmanschap M. Cost effectiveness of an active implementation strategy for the Dutch physiotherapy guideline for low back pain. *Health policy*. 2005 Dec;75(1):85-98.
13. Karrholm J, Ekholm K, Ekholm J, Bergroth A, Ekholm KS. Systematic co-operation between employer, occupational health service and social insurance office: a 6-year follow-up of vocational rehabilitation for people on sick-leave, including economic benefits. *J Rehabil Med*. 2008 Aug;40(8):628-36.
14. Smit F, Willemsse G, Koopmanschap M, Onrust S, Cuijpers P, Beekman A. Cost effectiveness of preventing depression in primary care patients: randomised trial. *The British journal of psychiatry : the journal of mental science*. 2006 Apr;188:330-6.
15. OECD. *Mental Health and Work: Netherlands*. 2014.
16. Loisel P, Anema JR. *Handbook of Work Disability. Prevention and Mangement*. New York: Springer Science+Business Media; 2013. 513 p.
17. van der Klink JJJ, Burdorf A, Schaufeli WB, vd. Wilt GJ, Zijlstra FRH, Brouwer S, et al. Duurzaam inzetbaar: werk als waarde [Sustainably employable: the value of work]. Den Haag: 2011.
18. Evers HGJ. *The Improved Gatekeeper Act. A practical guideline [In Dutch: De Wet Verbetering Poortwachter. Handleiding voor de praktijk]*. Deventer: Kluwer Juridisch; 2007.
19. Drummond M, Brandt A, Luce B, Rovira J. Standardizing methodologies for economic evaluation in health care. Practice, problems, and potential. *Int J Technol Assess Health Care*. 1993 Winter;9(1):26-36.
20. Tompa E, Verbeek J, van Tulder M, de Boer A. Developing guidelines for good practice in the economic evaluation of occupational safety and health interventions. *Scand J Work Environ Health*. 2010 Jun;36(4):313-8.
21. Drummond M, Manca A, Sculpher M. Increasing the generalizability of economic evaluations: recommendations for the design, analysis, and reporting of studies. *Int J Technol Assess Health Care*. 2005 Spring;21(2):165-71.

CHAPTER 1

22. Drummond M, McGuire A. *Economic Evaluation in Health Care: Merging theory with practice* New York: Oxford University Press; 2001.
23. Drummond M, Sculpher MJ. *Methods for the Economic Evaluation of Health Care Programmes*: Oxford University Press; 2005.
24. Tompa E, Culyer AJ, Dolinschi R. *Economic Evaluation of Interventions for Occupational Health and Safety: Developing Good Practice*. New York: Oxford University Press; 2008.
25. Young A. Return to Work Stakeholders' Perspectives on Work Disability. In: Loisel P, Anema JR, editors. *Handbook of Work Disability: Prevention and Management*. New York: Springer Science+Business Media; 2013.
26. Tompa E, Dolinschi R, de Oliveira C. Practice and potential of economic evaluation of workplace-based interventions for occupational health and safety. *J Occup Rehabil*. 2006 Sep;16(3):375-400.

PART I

METHODOLOGICAL CHALLENGES



CHAPTER

2

Quality appraisal of generic self-reported instruments measuring health-related productivity changes: A systematic review

Cindy Noben
Silvia Evers
Frans Nijhuis
Angelique de Rijk

Abstract

Background: Health impairments can result in disability and changed work productivity imposing considerable costs for the employee, employer and society as a whole. A large number of instruments exist to measure health-related productivity changes; however their methodological quality remains unclear. This systematic review critically appraised the measurement properties in generic self-reported instruments that measure health-related productivity changes to recommend appropriate instruments for use in occupational and economic health practice.

Methods: PubMed, PsycINFO, Econlit and Embase were systematically searched for studies whereof: (i) instruments measured health-related productivity changes; (ii) the aim was to evaluate instrument measurement properties; (iii) instruments were generic; (iv) ratings were self-reported; (v) full-texts were available. Next, methodological quality appraisal was based on COSMIN elements: (i) internal consistency; (ii) reliability; (iii) measurement error; (iv) content validity; (v) structural validity; (vi) hypotheses testing; (vii) cross-cultural validity; (viii) criterion validity; and (ix) responsiveness. Recommendations are based on evidence syntheses.

Results: This review included 25 articles assessing the reliability, validity and responsiveness of 15 different generic self-reported instruments measuring health-related productivity changes. Most studies evaluated criterion validity, none evaluated cross-cultural validity and information on measurement error is lacking. The Work Limitation Questionnaire (WLQ) was most frequently evaluated with moderate respectively strong positive evidence for content and structural validity and negative evidence for reliability, hypothesis testing and responsiveness. Less frequently evaluated, the Stanford Presenteeism Scale (SPS) showed strong positive evidence for internal consistency and structural validity, and moderate positive evidence for hypotheses testing and criterion validity. The Productivity and Disease Questionnaire (PRODISQ) yielded strong positive evidence for content validity, evidence for other properties is lacking. The other instruments resulted in mostly fair-to-poor quality ratings with limited evidence.

Conclusions: Decisions based on the content of the instrument, usage purpose, target country and population, and available evidence are recommended. Until high-quality studies are in place to accurately assess the measurement properties of the currently available instruments, the WLQ and, in a Dutch context, the PRODISQ are cautiously preferred based on its strong positive evidence for content validity. Based on its strong positive evidence for internal consistency and structural validity, the SPS is cautiously recommended.

Keywords: absenteeism, presenteeism, disability, productivity, self-report, psychometrics, validation.

Published as: C. Y. Noben, S. M. Evers, F. J. Nijhuis, A. E. de Rijk. Quality appraisal of generic self-reported instruments measuring health-related productivity changes: a systematic review. *BMC Public Health* 2014;14:115.

Introduction

When information about the costs of alternative treatments is to be used to guide healthcare policy decision making, it is the total budget needed to treat patients with the disease that is relevant. Estimates of these total costs are based on various cost categories, such as direct health care costs (costs of healthcare resources used by patients) and indirect healthcare costs (costs due to lost productivity). As it is known, health impairments among workers can result in considerable costs due to work disability, sickness absence, and productivity loss at work, all imposing a substantial financial burden for the employee, employer and society as a whole. Various studies on different diseases have shown that indirect costs, henceforth called 'productivity costs', contributed substantially to total costs, illustrating how important the consequences of disease are for work performance [1-4]. Productivity costs refer to the costs associated with lost or impaired ability to work or to engage in leisure activities due to morbidity and lost economic productivity due to death [5]. A study in the Netherlands showed that the productivity costs due to low back pain can be as high as 93% of the total costs of this impairment [3]. In Germany, productivity costs due to asthma amounted to 75% of the total costs [2]. A large study in the USA among workers with common health conditions showed that productivity costs substantially exceeded the direct costs. Moreover, presenteeism costs, or costs due to reduced productivity while still at work, appeared to represent up to 60% of all costs [4].

Converting the changes of health-related productivity into a financial metric makes these changes more interpretable. However, there is no agreement on how to quantify time lost due to health impairments or how to assign a monetary value to the lost productivity. To help improve the comparability and interpretability of productivity changes, a sound estimation of productivity costs requires sound measurements of the relevant components. The comparability of estimated productivity costs is hampered by substantial differences in the costs of the items considered and the methods used for measuring sickness absence and presenteeism, as well as differences in and insufficient methodology used in the valuation of these measurement tools.

In the last decades, a large number of measurement methods and instruments have been developed to quantify health-related productivity changes. These instruments are preferably self-administered by workers with health impairments because objective measurements of productivity changes are unable to capture reduced productivity while still at work (i.e. presenteeism). Several studies have shown that presenteeism contributes substantially to the estimated total costs of health impairments among workers [6-10]. The comparability across studies estimating productivity changes and associated costs is poor, since methods of measuring changed productivity seem to vary considerably.

There is thus an urgent need for practical and applicable knowledge and insight into the reliability, validity and responsiveness of these instruments. Regarding the validity of the instruments, one should keep in mind that the extent to which a valid measurement of productivity loss, especially presenteeism, can be achieved is often influenced by many factors (e.g. the amount of teamwork required in the job, the work setting, the desired actual production output, etc.) [6, 11].

Although several researchers have provided comprehensive reviews of existing instruments that measure productivity changes, the methodological quality of the reviewed studies remains unclear [12, 13]. Consequently, judgments on the quality of the studies cannot be made. If the methodological quality of a study on the measurement properties of a specific instrument is appropriate, the results can be used to assess the quality of the instrument at issue, including its measurement properties. However, if the methodological quality of the study is inadequate, the results cannot be trusted and the quality of the instrument under study remains unclear, despite of the magnitude or strength of the estimates presented [14]. Therefore, in this systematic review both the methodological quality of the study and the quality of the instrument, based on its psychometric properties, are taken into account. The main aim of this systematic review is therefore to critically appraise and compare the measurement properties of generic, self-reported instruments measuring health-related productivity changes.

Methods

Search strategy

A systematic review was conducted of studies evaluating the measurement properties of generic, self-reported health-related productivity instruments. In February 2013, the following relevant electronic databases were searched for English-language peer-reviewed journal articles: Medline (PubMed), PsychINFO (EBSCOhost), Embase and EconLit. The search query (see Additional file 1) consisted of a combination of related terms for the following features: the construct of interest (i.e. workplace productivity loss OR absenteeism OR presenteeism) 'AND' studies on measurement properties 'AND' the instrument of interest (i.e. generic AND self-report). The complete search strategies, including a sensitive search filter and exclusion filter [15] can be obtained via the corresponding author. No restrictions for the year of publication were made. Additional relevant studies were identified by performing database searches using the names of retrieved instruments and information from the retrieved reference lists.

Eligibility criteria

Study selection was based on the following eligibility criteria: (i) the aim of the study is the development or evaluation of the measurement properties of the instrument; (ii) the instrument under review measures health-related productivity changes; (iii) the instrument is generic and thus not solely focus on productivity changes due to a specific health impairment; (iv) health-related productivity change is rated from a worker's perspective (i.e. is self-reported); (v) full-text articles published in English were available.

Selection process

Three reviewers (CYGN, SMAAE and AER) independently determined the eligibility of all studies based on the title, keywords and abstracts. Studies in which an instrument of potential interest was used as an outcome measure, such as in intervention trials, were excluded. Review articles were excluded as well.

All reference lists and instruments of interest mentioned in all articles (both included and excluded from the review) were used in a secondary database search to identify additional relevant studies. If there was any doubt as to whether the studies met the eligibility criteria after individual selection, discrepancies were resolved through discussion between the reviewers, and a consensus decision was made. In case of remaining uncertainty, the full text was reviewed. Reasons for excluding the abstracts can be retrieved from the first author.

Measurement properties

The Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) covers relevant measurement properties for health-related, patient-reported outcomes and is based on international consensus [14, 16, 17].

The COSMIN checklist was used to rate the methodological quality of each article evaluating measurement properties. It covers three main quality domains: reliability, validity and responsiveness. Each domain is divided into different measurement properties and aspects with questions ranking the quality level of the design and statistical analyses [16]. Table 1 illustrates the most appropriate measures to critically appraise the measurement properties.

Table 1 - Description of the measurement domains, properties, aspects, and statistics and methods

Domains	Properties	Aspects	Statistics / Methods	
Reliability	Internal Consistency		Cronbach's alpha or Kuder-Richardson formula (KR-20) to determine relevance Factor analysis or principal component analysis to determine whether items form one or more than one scale	
	Reliability		Intraclass correlation coefficient (ICC) or Cohen's kappa	
Validity	Measurement error		Standard error of measurement (SEM) Smallest detectable change (SDC) Change beyond measurement error Limits of agreement (LoA) Minimal important change to determine the adequacy of measurement error	
	Content validity	Face validity	Assessment of relevance of all items for the construct, aim and target group Assessment of important missing items	
	Construct validity	Structural validity	Factor analysis to confirm the number of subscales present	
		Hypotheses testing	Assessment of a priori hypotheses, clearly indicating both direction and magnitude of the correlation or difference	
		Cross-cultural validity	Assessment of adequate reflection of the performance of the items of the original instrument	
	Criterion validity		Correlation Area under the receiver operator characteristics curve (AUC) Sensitivity and specificity	
	Responsiveness			Assessment of a priori hypotheses focusing on the change score of an instrument in the hypotheses Area under the receiver operator characteristic curve (AUC)

The domain ‘reliability’ is defined as the degree to which the measurement is free from error and the degree to which patient scores have not changed for repeated measurements over different sets of items from the same questionnaire (internal consistency), over time (test-retest), over different persons (inter-rater), or over different occasions (intra-rater). Reliability is further assessed based on the measurement properties: (i) internal consistency: the degree of interrelatedness among the items; (ii) reliability: the proportion of total variance in the measurement due to ‘true’ (free of error) differences between patients; and (iii) measurement error: systematic and random error of patient scores (not attributed to true changes in the construct to be measured).

The second domain, ‘validity’, is described as the degree to which the instrument measures what it purports to measure. Three measurement properties are assessed: (i) content validity: the degree to which content of the instrument adequately reflects the construct to be measured, including face validity; (ii) construct validity, or the degree to which the scores of the instrument are consistent with the hypotheses, is divided into three aspects: (a) structural validity: the degree to which the scores of the instrument are an adequate reflection of the dimensionality of the construct to be measured; (b) hypotheses testing: the degree to which scores of the instrument are consistent with hypotheses; (c) cross-cultural validity: the degree to which performance of the items on a translated or culturally adapted instrument are an adequate reflection of the performance of the items of the original version of the instrument; and finally (iii) criterion validity: the degree to which the scores of the instrument are an adequate reflection of a ‘gold standard’ [17].

The third domain, ‘responsiveness’, is defined as the ability of the instrument to detect change over time in the construct to be measured.

Data extraction

Three reviewers (CYGN, AER, SMAAE) independently extracted the target data from the full text articles. The studies (n=25) are described based on information retrieved from the original publications and contain information regarding the study country, population, sampling methods, setting, age, gender and response rates. The instruments (n=15) are described based on information retrieved from the original publications (e.g. content, number of items, rating, item example, recall, and discipline).

Methodological quality assessment

To determine the methodological quality of the studies, each was assessed independently by three reviewers (CYGN, AER, SMAAE). Consensus was reached by pairing the reviewers’ results. The pairing was CYGN and SMAAE (pair 1) and CYGN and AER (pair 2). When a pair of reviewers disagreed, consensus was reached through discussion within the project group (CYGN, SMAAE, FJN, AER). The meth-

odological quality assessment of the studies was conducted by scoring each of the nine measurement properties as presented in nine boxes by the COSMIN-checklist. A four-point rating scale was used (i.e. poor, fair, good, or excellent) to calculate the quality score per measurement property. The scores took into account, for example, the used sample size which may differ between methods (rules of thumbs for factor analyses vary between a subject-to-variables ratio of 4:1 to 10:1 or a sample size of 50 is needed to obtain confidence intervals from 0.70-0.90 around an ICC to assess reliability estimates); the missing items and the missing responses per items and how they were handled; the description of the comparator; and the appropriateness of the statistics (e.g. the internal consistency statistic only gets an interpretable meaning when the interrelatedness among the items is determined of a set of items that together form a reflective model). The methodological quality was determined per study for each measurement property separately by taking the lowest rating of the items to that measurement property in each box (worse score counts) [14].

Best evidence synthesis

For each instrument, a best evidence synthesis was performed by combining the methodological quality score for each measurement property per instrument as assessed in this study (excellent, good, fair or poor) with the consistency of their results concerning the measurement property (positive or negative evidence for a measurement property). The ratings correspond to the Cochrane Back Review Group levels of evidence [18]. The levels of evidence are strong, moderate, limited, conflicting, or unknown. A strong level of evidence represents consistent findings in multiple studies of good methodological quality *or* in one study of excellent methodological quality; moderate levels of evidence represent consistent findings in multiple studies of fair methodological quality *or* in one study of good methodological quality; a limited level of evidence occurs when one study of fair methodological quality is being presented; an unknown level of evidence is noted when only studies of poor methodological quality are obtained and a conflicting level of evidence represents for multiple studies with conflicting findings.

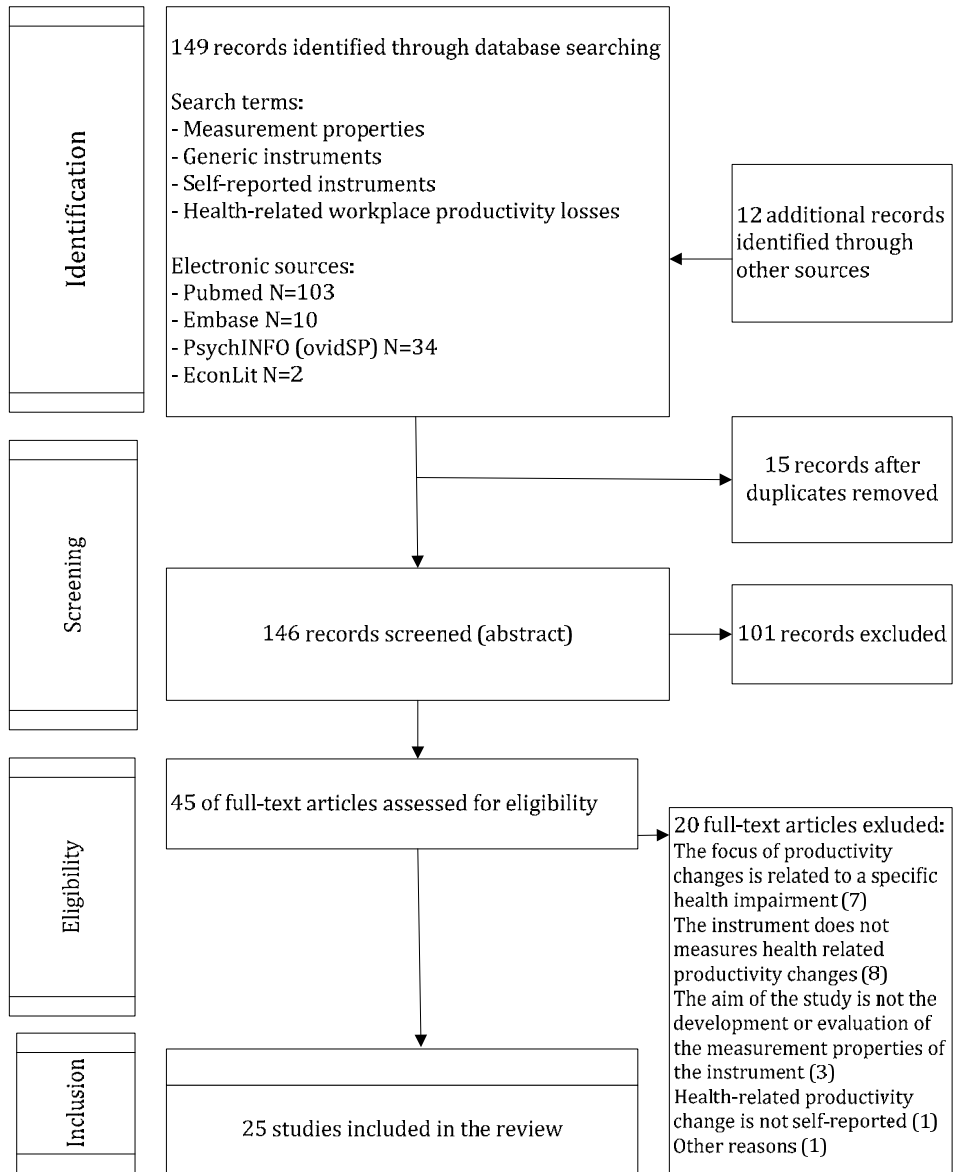
Results

Study selection

The first search in MEDLINE (PubMed) resulted in 103 hits. The searches in PsychINFO, Embase and EconLit resulted in 34, 10 and 2 hits, respectively. Manual searches based on reference lists and names of original instruments identified 12 additional records. Duplicates (N=15) were automatically discarded via Endnote X4

resulting in 146 records to be screened. After screening abstracts and titles, 45 articles were reviewed, and 25 met the eligibility criteria. The most common reasons for exclusion were: (i) the main aim of the study was not about assessing the psychometric properties of a health-related productivity instrument; (ii) the focus of productivity changes was related to specific health impairment; (iii) the instrument did not measure health-related productivity changes; (iv) health-related productivity was not self-reported. A flow chart illustrates the process of inclusion (see Figure 1).

Figure 1. Flowchart of inclusion



General description of the identified studies

Finally, 25 articles evaluating 15 different instruments were included in this study. The characteristics of the 25 included studies are presented in Table 2. Several studies evaluated the measurement properties of multiple instruments and are

therefore mentioned several times [19-21]. Two studies were conducted in the UK and Ireland [22, 23], three in the Netherlands [20, 24, 25], three in Canada [26-28], and 17 in the USA [19, 21, 29-43]. The largest proportion (44%) of the studies focused on physical disorders [20, 22, 23, 25-28, 30, 35, 37, 42]. Mental health problems were only addressed in two studies [19, 29]. Populations with varying health problems (physical and mental constraints) were included in four studies [24, 31, 36, 41]. Finally, workers with non-specified health issues [32, 34, 38, 40, 43] or workers with job-related injuries [21, 33, 39] were also included in five and three studies respectively. Twelve studies were conducted at the workplace [20, 21, 24, 31-34, 38-41, 43] and nine studies were conducted in a healthcare institution [19, 26-29, 35-37, 39]. One study was conducted at the workplace and at a healthcare institution and is therefore mentioned twice [39]. One study took place at an occupational medicine college [30], and four studies derived their participants from a database [22, 23, 25, 42].

Table 2. Characteristics of the identified studies

Study	Country	Instru- ment	Population	Sampling	Setting	Mean age (SD)	Gender (% female)	Response rate (%)
Beaton [26]	Canada	WLQ	Musculoskeletal disorders	Convenience	Health care	42.5 (10.1)	54%	33%
Endicott [29]	USA	EWPS	Major depressive disorders	Convenience	Health care	41 (9.6)	70%	N.A.
Erickson [19]	USA	WLQ WPSI EWPS WPPI	Anxiety disorders (i) minimal-to-mild (ii) moderate-to-severe	Convenience	Health care	(i) 37.5 (12.2) (ii) 34.2 (9.8)	(i) 48.8% (ii) 75%	51%
Forst [30]	USA	AMA-guide	Low back injuries	Convenience	Academy	N.A.	N.A.	100%
Goetzel [31]	USA	WPSI	Allergies, respiratory infections, arthritis*	N.A.	Workplace	47 (N.A.)	30%	N.A.
Kessler [33]	USA	WHO HPQ	Job-related accidents-injuries (i) airline reservation agents (ii) customer service reps. (iii) executives (iv) railroad engineers	Convenience	Workplace	(i) 30-44 (2.4) (ii) 30-44 (2.2) (iii) 45-59 (2.0) (iv) 45-59 (1.7)	(i) 80.3% (ii) 47.2% (iii) 19.3% (iv) 2.4%	(i) 39% (ii) 29% (iii) 50% (iv) 57%
Kessler [32]	USA	WHO HPQ	Non-specified health status	Convenience	Workplace	N.A.	N.A.	N.A.
Koopman [34]	USA	SPS	Non-specified health status	Convenience	Workplace	46.5 (9.4)	47.9%	74%
Koopman-schap [24]	NL	PRODISQ	Musculoskeletal complaints, stress factors, back complaints**	Convenience	Workplace	N.A.	N.A.	N.A.

Study	Country	Instru- ment	Population	Sampling	Setting	Mean age (SD)	Gender(% female)	Response rate (%)
Kumar [35]	USA	HRPQ-D	Infectious Mononucleosis	Convenience	Health care	19.3 (3.6)	41.9%	97%
Lerner [36]	USA	WLQ	RA, chronic daily headache syn- drome and epilepsy	Convenience	Health care	41.3 (11.1)	72.7%	N.A.
Lerner [37]	USA	WLQ	Osteoarthritis	Convenience	Health care	53.7 (7.1)	65.4%	30%
Meerding [20]	NL	HLQ Q&Q	Musculoskeletal complaints (i) industrial workers (ii) construc- tion workers	Convenience	Work- place	(i) 35 (N.A.) (ii) 42 (N.A.)	(i) 0%(ii) 6%	(i) 69% (ii) 85%
Ozminkan- ski [21]	USA	WLQ WPSI	Job-related accidents-injuries	Consecutive	Work- place	37.77 (N.A.)	34.31%	48%
Prochaska [38]	USA	WBA-P	Non-specified health status	Convenience	Work- place	47.6 (11.4)	56.8%	N.A.
Reilly [39]	USA	WPAI	Job-related accidents-injuries (i) Group 1 (ii) Group 2	Convenience	Health care / Work- place	(i) 37.2 (9.7) (ii) 39.4 (11.8)	(i) 5.5% (ii) 60.8%	48%
Roy [28]	Canada	WLQ	Chronic upper-extremity disor- ders	Convenience	Health care	46 (9)	53%	83%
Shikiar [40]	USA	HWQ	Non-specified health status	Convenience	Work- place	N.A.	N.A.	45%
Stewart [43]	USA	WHI	Non-specified health status	Convenience	Work- place	N.A.	66%	72%

Study	Country	Instru- ment	Population	Sampling	Setting	Mean age (SD)	Gender(% female)	Response rate (%)
Tang [27]	Canada	WLQ	Shoulder and elbow disorders	Convenience	Health care	43.3 (11.5)	42.5%	87%
Turpin [41]	USA	SPS	Allergies, arthritis or joint pain/stiffness, asthma ***	Convenience	Work- place	43.2 (N.A.)	29.4%	63%
Van Rooijen [25]	NL	HLQ	(i) Migraine (ii) spinal cord injury (iii) knee injury (iv) hip injury	Convenience	Data base	N.A.	N.A.	(i) 58% (ii) 81% (iii) 76% (iv) 75%
Walker [42]	USA	WLQ	Rheumatoid arthritis	Continuously	NDB	53.5 (10.5)	77%	98%
Zhang [22]	UK & Ireland	WPAl	Rheumatoid arthritis	Convenience	ERAN	52.1 (10)	72%	53%
Zhang [23]	UK & Ireland	VOLP	Rheumatoid arthritis	Convenience	ERAN	52 (10)	72%	81%

WLQ = Work Limitations Questionnaire; EWPS = Endicott Work Productivity Scale; N.A. = Not Available; WPSI = Work Productivity Short Inventory; WPAl = Work Productivity and Activity Impairment Instrument; AMA-guide = American Medical Association-Guides; WHO HPQ = World Health Organization Health and work Performance Questionnaire; SPS = Stanford Presenteeism Scale; PRODISQ = PROductivity and Disease Questionnaire; HRPQ-D = Health-Related Productivity Questionnaire Diary; HLQ = Health and Labor Questionnaire; Q&Q = Quality and Quantity questionnaire; WBA-P = Well-Being Assessment for Productivity; HWQ = Health and Work Questionnaire; WHI = Work and Health Interview; VOLP = Valuation of Lost Productivity questionnaire; NDB = National Databank for Rheumatic Diseases ; ERAN = The Early Rheumatoid Arthritis Network * Continued: asthma, anxiety disorders, depression, bipolar disorder, stress, diabetes, hypertension, migraine, other major headaches, coronary heart disease/high cholesterol ** Continued: dyspepsia, chronic fatigue syndrome, psoriasis, rheumatoid arthritis *** Continued: back or neck disorder, breathing disorder, depression, anxiety or emotional disorder, diabetes; hearth or circulatory problem, migraine/chronic headaches, stomach or bowel disorder

Fifteen different generic self-reported instruments measuring health-related productivity changes are included in this study. The general characteristics of the instruments are presented below. An additional Table shows a more detailed description of the instruments (see Additional file 2).

In Table 3, the methodological quality of each study per measurement property and instrument is presented. The Work Limitations Questionnaire (WLQ) is the most frequently evaluated instrument [19, 21, 26-28, 36, 37, 42], followed by the Work Productivity and Activity Impairment Instrument (WPAI) [19, 22, 39], and the Work Productivity Short Inventory (WPSI) [19, 21, 31]. Criterion validity was evaluated in 21 studies [19-23, 25, 28-41, 43]. Hypotheses testing was evaluated in 15 studies [19, 22, 26-36, 39, 41]. Only one study addressed measurement error [39], and none of the studies evaluated cross-cultural validity.

Table 3. Methodological quality of each study per measurement property per instrument

Instrument and paper	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypotheses testing	Cross-cultural validity	Criterion validity	Responsiveness
WLQ									
Beaton [26]	Poor	Fair	.	.	Fair
Erickson [19]	Poor	Poor	.	Poor	Poor
Lerner [36]	Poor	Fair	.	Good	.	Poor	.	Poor	.
Lerner [37]	Good	.	.	Fair	Good	.	.	Poor	.
Ozminkowski [21]	Fair	.
Roy [28]	Fair	.	Good	Fair
Tang [27]	Fair	.	.	Fair	.	Fair	.	.	.
Walker [42]	Excellent	.	.	.	Good
EWPS									
Endicott [29]	Poor	Poor	.	.	.	Fair	.	Poor	.
Erickson [19]	Poor	Poor	.	Poor	Poor
WPAI									
Erickson [19]	Poor	Poor	.	Poor	Poor
Reilly [39]	.	Poor	Poor	.	.	Fair	.	Fair	.
Zhang [22]	Fair	.	Fair	.
WPSI									
Erickson [19]	Poor	Poor	.	Poor	Poor
Goetzel [31]	.	Poor	.	Poor	.	Poor	.	Poor	Poor
Ozminkowski [21]	Fair	.
AMA-guides									
Forst [30]	.	Poor	.	.	.	Poor	.	Poor	.
WHO-HPQ									
Kessler 2003 [33]	.	.	.	Good	.	Poor	.	Good	.

Instrument and paper	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypotheses testing	Cross-cultural validity	Criterion validity	Responsiveness
Kessler [32]	.	Fair	.	Good	Fair	Fair	.	Fair	Poor
SPS									
Koopman [34]	Good	.	.	Fair	Good	Fair	.	Fair	.
Turpin [41]	Good	Fair	.	.	Good	Fair	.	Fair	.
PRODISQ									
Koopmanschap [24]	.	.	.	Excellent
HRPQ-D									
Kumar [35]	.	Fair	.	Poor	.	Poor	.	Fair	Fair
HLQ									
Meerding [20]	.	Fair	Fair	.
van Rooijen [25]	Fair	.
Q&Q									
Meerding [20]	.	Fair	Fair	.
WBA-P									
Prochaska [38]	Good	.	.	Good	.
HWQ									
Shiklar [40]	Fair	.	.	.	Fair	.	.	Fair	Poor
WHI									
Stewart [43]	.	Good	Good	Fair
VOLP									
Zhang [23]	.	Good	Good	.

WLQ = Work Limitations Questionnaire; EWPS = Endicott Work Productivity Scale; WPSI = Work Productivity Short Inventory; WPAI = Work Productivity and Activity Impairment Instrument; AMA-guide = American Medical Association-Guides; WHO HPQ = World Health Organization Health and work Performance Questionnaire; SPS = Stanford Presenteeism Scale; PRODISQ = Productivity and Disease Questionnaire; HRPQ-D = Health-Related Productivity Questionnaire Diary; HLQ = Health and Labor Questionnaire; Q&Q = Quality and Quantity questionnaire; WBA-P = Well-Being Assessment for Productivity; HWQ = Health and Work Questionnaire; WHI = Work and Health Interview; VOLP = Valuation of Lost Productivity questionnaire

CHAPTER 2

Table 4 presents the level of evidence for each measurement properties per instrument by synthesizing the results per instrument and their accompanying level of evidence. The methodological quality for each instrument and measurement property according to the three main quality domains (reliability, validity and responsiveness) is outlined below, along with a brief description of the aim and the content of the instrument. The synthesis of results per instrument and their accompanying level of evidence are presented in line with the content of Table 4.

The evidence for each measurement property is mostly limited; for eight out of 15 instruments at least 50% of the information on measurement properties is lacking.

Table 4. Level of evidence per measurement property per instrument

Instrument and paper	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypotheses testing	Cross-cultural validity	Criterion validity	Responsiveness
WLQ [19, 21, 26-28, 36, 37, 42]	+/-	-	N.A.	++	+++	--	N.A.	+/-	--
EWPS [19, 29]	?	?	N.A.	N.A.	N.A.	+/-	N.A.	?	?
WPAI [19, 22, 39]	?	?	?	N.A.	N.A.	--	N.A.	--	?
WPSI [19, 21, 31]	?	?	N.A.	?	N.A.	?	N.A.	+/-	?
AMA-guides [30]	N.A.	?	N.A.	N.A.	N.A.	?	N.A.	?	N.A.
WHO-HPQ [32, 33]	N.A.	-	N.A.	+++	+	+/-	N.A.	+/-	?
SPS [34, 41]	+++	-	N.A.	-	+++	++	N.A.	++	N.A.
PRODISQ [24]	N.A.	N.A.	N.A.	+++	N.A.	N.A.	N.A.	N.A.	N.A.
HRPQ-D [35]	N.A.	-	N.A.	?	N.A.	?	N.A.	-	-
HLQ [20, 25]	N.A.	-	N.A.	N.A.	N.A.	N.A.	N.A.	--	N.A.
Q&Q [20]	N.A.	-	N.A.	N.A.	N.A.	N.A.	N.A.	-	N.A.
WBA-P [38]	N.A.	N.A.	N.A.	N.A.	+	N.A.	N.A.	+	N.A.

Instrument and paper	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypotheses testing	Cross-cultural validity	Criterion validity	Responsiveness
HWQ [40]	+	N.A.	N.A.	N.A.	-	N.A.	N.A.	-	?
WHI [43]	N.A.	-	N.A.	N.A.	N.A.	N.A.	N.A.	-	-
VOLP [23]	N.A.	+	N.A.	N.A.	N.A.	N.A.	N.A.	-	N.A.

WLQ = Work Limitations Questionnaire; EWPS = Endicott Work Productivity Scale; WPSI = Work Productivity Short Inventory; WPAI = Work Productivity and Activity Impairment Instrument; AMA-guide = American Medical Association-Guides; WHO HPQ = World Health Organization Health and work Performance Questionnaire; SPS = Stanford Presenteeism Scale; PRODISQ = Productivity and Disease Questionnaire; HRPQ-D = Health-Related Productivity Questionnaire Diary; HLQ = Health and Labor Questionnaire; Q&Q = Quality and Quantity questionnaire; WBA-P = Well-Being Assessment for Productivity; HWQ = Health and Work Questionnaire; WHI = Work and Health Interview; VOLP = Valuation of Lost Productivity questionnaire N.A. = Not Available; +++ = Strong positive level of evidence; ++ = Moderate positive level of evidence; + = Limited positive level of evidence; - = Limited negative level of evidence; -- = Moderate negative level of evidence; +/- = Conflicting level of evidence; ? = Unknown level of evidence due to poor methodological quality

The Work Limitations Questionnaire (WLQ)

The WLQ measures the degree to which health problems interfere with specific aspects of job performance and the productivity impact of these work limitations. The questionnaire consists of four domains and includes a total of 25 items to be rated on a five-point scale [19, 21, 26-28, 36, 37, 42].

Reliability: Six studies evaluated the internal consistency of the WLQ [19, 26, 27, 36, 37, 42]. The methodological quality of three studies is poor due to small sample sizes [19, 26] and lacking factor analysis to check the unidimensionality of the instrument scale [19, 36]. One article is of good methodological quality regarding internal consistency [37]. Although it was not explicitly described, it was possible to deduce how missing items were handled. Furthermore, the authors referred to another study in which factor analysis was performed in a similar population, making it possible to check the unidimensionality of the scale. The paper of Tang, Pitts, Solway & Beaton is of fair methodological quality due to the moderate sample size, and it was unclear whether the factor analysis referred to was appropriate to assess the internal consistency of the WLQ because it was not conducted in a similar population [27]. The last study that evaluated the internal consistency of the WLQ was conducted by Walker, Michaud & Wolfe which is of excellent methodological quality [42]. Sample sizes were adequate, the article clearly described how missing items were handled and the unidimensionality of the scale was checked appropriately. The internal consistency statistic was calculated for each separate WLQ component. Because the findings are conflicting (three studies with poor methodological quality, one of fair, one of good, and one of excellent methodological quality and conflicting findings ranging from moderate ($C\alpha$ 0.74) to very high ($C\alpha$ 0.97) Cronbach's alphas) evidence synthesis of the WLQ resulted in conflicting evidence for internal consistency.

Only one study assessed the reliability [36], which is of fair methodological quality regarding this psychometric property. Although the authors calculated intraclass correlation coefficients (ICC 0.58 – 0.92) and a weighted kappa for continuous and dichotomous scores respectively, it was unclear if patients were stable in the interim period of the measurement. Furthermore, the time interval was not stated, the sample size was moderate and there was no description of how missing items were handled [36]. Evidence synthesis of the WLQ resulted in limited negative evidence for reliability (fair methodological quality and correlation coefficients <0.70).

Validity: Three studies examined the content validity [27, 36, 37]. Lerner et al. [36] reported high validity after assessing whether all items referred to relevant measurement constructs, whether all items were relevant for the study population, and whether all items together comprehensively reflected the construct. The methodological quality of the study is good. Both other studies are of fair methodological quality because the aspects of the construct to be measured were poorly described, and it was not taken into account whether all items referred to relevant aspects of

the WLQ [27, 37]. Evidence synthesis of the WLQ resulted in moderate positive evidence for content validity.

The structural validity was assessed in two studies [37, 42], resulting in good methodological quality. Both studies performed confirmatory factor analysis, which was an appropriate analysis in view of the existing information. Further, the sample size included was adequate, and the way the missing items were handled was either described or could be deduced [37, 42]. Evidence synthesis of the WLQ resulted in strong positive evidence for structural validity due to the consistent positive findings in two studies of good methodological quality.

Five studies performed hypotheses testing [19, 26-28, 36]. Three studies formulated a priori hypotheses. However, the measurement properties of the comparator instrument were not adequately described, and there was no evidence that the comparator could be applied to a similar study population [26-28]. Therefore, the methodological quality regarding hypotheses testing is rated fair in these three studies. Two studies are of poor methodological quality because they did not state clear a priori hypotheses [19, 36]. Due to the lack of information on the measurement properties of the comparator instruments, it was unclear what was expected in one study [36]. Evidence synthesis of the WLQ resulted in moderate negative evidence for hypotheses testing (three studies with fair methodological quality and two with poor methodological quality and only 60% of the results were in accordance with the hypotheses).

Criterion validity was evaluated in five studies [19, 21, 28, 36, 37]. Three of the studies are rated as having poor methodological quality regarding criterion validity [19, 36, 37]. In two studies it was not clear whether the criterion used could be considered an adequate gold standard [19, 36]. In all three studies, the statistical methods applied were inappropriate to assess criterion validity. Although claiming acceptable criterion validity, two studies failed to determine the calculation of specificity and sensitivity for dichotomous scores and solely mentioned the sensitivity to change (in clinical improvements) related to productivity [19, 36]. Lerner, Reed, Massarotti, Wester & Burke [37] did not calculate correlations, or areas under the receiver operating curves, for continuous scores but merely calculated averages. Although Ozminkowski, Goetzl, Chang & Long [21] assumed the criterion used could be considered as a gold standard, no evidence was provided. Therefore, the methodological quality is rated fair. One study [28] reported sufficient evidence on the criterion used to consider an adequate gold standard and applied appropriate statistical methods to assess criterion validity. Therefore the paper is of good methodological quality. Evidence synthesis of the WLQ resulted in conflicting evidence for criterion validity because three studies resulted in poor, one in fair, and one in good methodological quality and not all studies present the degree to which the scores of the instrument are an adequate reflection of a 'gold standard'.

Responsiveness: The responsiveness evaluated by two studies [26, 28] is of fair methodological quality because of a poor description of the constructs measured by the comparator instrument [28], a lack of information regarding measurement

properties of the comparator instrument [26, 28], a vague description of the hypotheses [26] and a moderate sample size [26]. Roy et al. [28] observed low correlation between change scores ($0.22 < r < 0.41$), demonstrating moderate responsiveness. To evaluate responsiveness, Erickson et al. [19] calculated the effect sizes (which were moderate, ranging from 0.36 to 0.73) between change scores based on changes in disease severity. The methodological quality is poor because the statistical method applied is not appropriate to test the hypotheses. Furthermore, the calculation methods for sensitivity and specificity were not determined, but were solely mentioned [19]. Evidence synthesis of the WLQ resulted in moderate negative evidence regarding responsiveness (fair and poor methodological quality and negative results).

There were no methodologically sound studies evaluating measurement error and cross-cultural validity of the WLQ.

The Endicott Work Productivity Scale (EWPS)

The EWPS assesses the degree to which a wide variety of mental and medical disorders of persons working in a wide variety of job settings, including self-employment, affect the work functioning of these persons. The EWPS contains 25 items scored on a five-point scale [19, 29].

Reliability: Internal consistency was studied in two papers that evaluated the measurement properties of the EWPS [19, 29]. Both studies are of poor methodological quality because no factor analyses were performed, and sample sizes were small [19, 29]. Although the internal consistency estimates show positive results with a Cronbach's alpha of 0.95 [19] and an internal consistency coefficient of 0.93 [29], evidence synthesis of the EWPS resulted in unknown evidence for internal consistency because only studies of poor methodological quality were available.

Endicott et al. [29] assessed the intraclass correlation coefficient of reliability (total score was 0.92). However, test-retest reliability was only assessed in a limited population including a small sample size ($N=16$), resulting in poor methodological quality for the paper [29]. Evidence synthesis of the EWPS resulted in unknown evidence regarding reliability because of the paper's poor methodological quality.

Validity : Hypotheses testing were performed in both papers, though no hypotheses were formulated a priori. However, in Endicott et al. [29] it was possible to deduce what was expected. Although the EWPS showed considerable promise as a sensitive measure for assessing the effects on work performance of various disorders, the instruments and constructs used were poorly described, and, therefore the paper is of fair methodological quality. Erickson et al. [19] confirmed their hypothesized expectations. However, the study is of poor methodological quality because expected differences and directions, and the magnitude of the differences were not stated, making it unclear what was expected. Evidence synthesis resulted in conflicting levels of evidence regarding hypotheses testing (one paper of fair, one of poor methodological quality, and conflicting findings in accordance to the deduced hypotheses).

Both papers are of poor methodological quality for criterion validity because the degrees to which the scores of instruments are an adequate reflection of a reasonable gold standard are unknown in both studies [19, 29]. Evidence synthesis of the EWPS resulted in unknown evidence for criterion validity because both studies were of poor methodological quality.

Responsiveness: Evidence synthesis of the EWPS resulted in unknown evidence for responsiveness because Erickson et al. [19] calculated the effect sizes between change scores based on changes in disease severity (-0.45). The methodological quality is poor because the statistical method applied is not appropriate to test the hypotheses.

There were no methodologically sound studies evaluating measurement error, content validity, structural validity or cross-cultural validity of the EWPS.

The Work Productivity and Activity Impairment Instrument (WPAI)

The WPAI measures the effect of general health and symptom severity on work productivity via six questions to be rated on a five-point scale [19, 22, 39].

Reliability: The evidence synthesis of the WPAI resulted in unknown evidence for internal consistency because Erickson et al. [19] did not apply factor analysis to check for unidimensionality of the instrument scale, and because the sample size was small, resulting in poor methodological quality.

Reliability was studied by Reilly, Zbrozek & Dukes [39]. Although correlation coefficients were calculated (ranging from 0.69 up to 0.95), the measurements were not independent, the time interval was not appropriate, and the test conditions were not similar. Therefore, the paper is of poor methodological quality. The evidence synthesis of the WPAI resulted in unknown evidence for reliability because of the paper's poor methodological quality.

The study by Reilly, Zbrozek & Dukes [39] included measurement error but because the measurements were not independent and the time interval was not appropriate, the methodological quality is poor. The evidence synthesis of the WPAI resulted in unknown evidence for measurement error due to the study's poor methodological quality.

Validity: Three studies performed hypotheses testing [19, 22, 39]. Although two studies [22, 39] stated a priori hypotheses and expected directions of the differences, the relationships were not always as pronounced as were expected and in both studies only a reference to a study on measurement properties of the comparator instrument was provided, resulting in fair methodological quality. Although it could be deduced what was expected in one study [19], and they confirmed their expectations, the study is of poor methodological quality because expected differences and directions, and the magnitude of the differences were lacking. This study is therefore of poor methodological quality. The evidence synthesis of the WPAI resulted in moderate negative evidence for hypotheses testing (two studies of fair methodological quality and one of poor methodological quality).

Criterion validity was assessed in all papers. The paper of Erickson et al. [19] used other instruments (EWPS and WLQ) as a comparator (significant correlation), which cannot be considered as a reasonable gold standard because the WPAI uses single item scales. Therefore, the paper is of poor methodological quality [19]. The two other papers [22, 39] are of fair methodological quality because it was not clear how missing items were handled [22] or whether the criterion used could be considered an adequate gold standard [39]. Correlations between the WPAI and SF-36 measures ranged from 0.20 to 0.52 [39] and from 0.34 up to 0.77 when comparing the WPAI outcomes with health status outcomes from another instrument [22]. The evidence synthesis of the WPAI regarding criterion validity resulted in moderate negative evidence (two studies of fair methodological quality and one of poor methodological quality and correlations <0.70).

Responsiveness: Responsiveness over time was evaluated by Erickson et al. [19] by calculating effect sizes (ranging from 0.19 to -0.87) between change scores based on changes in disease severity. The study results therefore in poor methodological quality. Evidence synthesis of the WPAI results in unknown evidence for responsiveness because of the study's poor methodological quality.

There were no methodologically sound studies evaluating the content validity, structural validity or cross-cultural validity of the WPAI.

The Work Productivity Short Inventory (WPSI)

The WPSI assesses the prevalence of medical problems that might influence work productivity based on 22 open questions [19, 21, 31].

Reliability: The internal consistency was studied by Erickson et al. [19], resulting in a Cronbach's alpha of 0.82. The paper results in poor methodological quality because no factor analyses were performed and sample sizes were small. Evidence synthesis of the WPSI resulted in unknown evidence for internal consistency because of the paper's poor methodological quality.

Goetzel, Ozminkowski & Long [31] evaluated the reliability of the WPAI but used only one measurement and calculated solely percentage agreement. Therefore, the paper is of poor methodological quality. Evidence synthesis of the WPSI resulted in unknown evidence for reliability because of the study's poor methodological quality.

Validity: The content validity was assessed by Goetzel, Ozminkowski & Long [31], and the paper results in poor methodological quality because there was no information on the degree to which the content of the instrument is an adequate representative of the constructs to be measured. No assessment of whether all items were relevant for the study population occurred, solely a simple description on how the most prevalent conditions in a firm are presented and allow for valid data collection. Evidence synthesis of the WPSI resulted in unknown evidence for content validity because of the study's poor methodological quality.

Two studies performed hypotheses testing [19, 31]. Erickson et al. [19] did not report expected differences and directions, and the magnitude of the differences

were not stated, making it unclear what was expected. Goetzel, Ozminkowski & Long [31] did not provide any information on the psychometrics of the comparator instrument and the statistical methods (t-test) were inadequate to test the hypotheses. Both papers result in poor methodological quality. Evidence synthesis of the WPSI resulted in unknown evidence regarding hypotheses testing because of the poor methodological quality of the studies.

Three studies assessed the criterion validity for the WPSI [19, 21, 31]. The paper of Erickson et al. [19] used a comparator without appropriate evidence that it could be considered a reasonable gold standard. Furthermore, only significance levels were reported, an inadequate method to assess criterion validity. The study of Goetzel, Ozminkowski & Long [31] compared three different versions, varying by recall period, based on coefficient of variation to assess criterion validity (resulting all higher than expected (>10)). As the validation of the WPSI cannot (yet) be empirically confirmed, the method used in this study is inappropriate. Furthermore, no correlation coefficients or area under the receiver operating curve were calculated [31]. Therefore, both papers are of poor methodological quality. Ozminkowski et al. [21] provided some information on the criterion used as a gold standard, but no evidence was provided. All correlations were positive and significantly different from zero, although none exceeded 0.38 in magnitude [21]. The methodological quality is rated fair. Evidence synthesis of the WPSI resulted in conflicting evidence for criterion validity.

Responsiveness: The responsiveness was evaluated by an effect size (0.49) between change scores based on changes in disease severity [19] and by significant marginal differences between groups with different conditions [31] making the statistical methods applied inappropriate for the hypotheses to be tested. Therefore, the paper results in poor methodological quality. Evidence synthesis of the WPSI resulted in unknown evidence regarding responsiveness due to the paper's poor methodological quality.

There were no methodologically sound studies evaluating measurement error, structural validity or cross-cultural validity of the WPSI.

The American Medical Association (AMA)-guides

The AMA-guides rate loss of functioning and evaluate workability after injury or illness via fifteen questions rated on a ten-point scale [30].

Reliability: Forst, Friedman & Chukwu [30] presents intra-class correlation coefficients comparing impairment raters using the fifth versus the sixth editions of AMA's-Guides and ranged from 0.65 to 0.77. The paper is of poor methodological quality because the sample size was small (N=16) and only reliability for each or between two editions was calculated. Evidence synthesis of the AMA-guides resulted in unknown evidence for reliability because of the paper's poor methodological quality.

Validity: Although hypotheses were vague, it was possible to deduce what was expected. However, because the sample size was small and other flaws in the design were presented; e.g. no information was provided regarding the measurement properties of the comparator instrument and the data presented a comparison of two versions of the same instrument, resulted in poor methodological quality of the paper of Forst, Friedman & Chukwu [30] is poor. Evidence synthesis of the AMA-guides therefore resulted in unknown evidence for hypotheses testing because of the study's poor methodological quality.

Because of the small sample size and the lack of psychometric characteristics of the previous version of the guide, the criterion used cannot be considered an adequate gold standard. Therefore, the methodological quality of the study by Forst, Friedman & Chukwu [30] is poor. Evidence synthesis of the AMA-guides resulted in unknown evidence for criterion validity because of the study's poor methodological quality.

There were no methodologically sound studies evaluating the internal consistency, measurement error, content validity, structural validity, cross-cultural validity or responsiveness of the AMA-guides.

The WHO Health and work Performance Questionnaire (WHO HPQ)

The WHO HPQ uses open questions, divided into three categories, to assess indirect workplace costs of illness by measuring absenteeism and presenteeism and critical incidents [32, 33].

Reliability: Kessler et al. [32] investigated the reliability (correlation 0.521) but did not report how missing items were handled, did not state the time interval, and did not clearly indicate whether the population was stable during the period of measurement. Therefore, the paper is of fair methodological quality. Evidence synthesis of the WHO HPQ resulted in limited negative evidence for reliability because of the study's fair methodological quality and correlation <0.70.

Validity: Two studies assessed the content validity of the WHO HPQ [32, 33]. Although one study did not clearly state the sample size [33], it was possible to assume that all items were relevant for the study population and that the purpose was to reflect on the measurement construct in both studies. Both papers are of good methodological quality. Evidence synthesis of the WHO HPQ resulted in strong positive evidence for content validity (good methodological quality and positive results).

To assess structural validity, Kessler et al. [32] conducted exploratory factor analysis. The overall model fit was excellent ($X^2 = 1.1, P = 0.3$). However, because it was not clear how missing items were handled, the study is of fair methodological quality. Evidence synthesis of the WHO HPQ resulted in limited positive evidence for structural validity.

Hypotheses testing were performed in both studies [32, 33]. In both studies hypotheses were vaguely formulated; however, it was possible to deduce what was expected. In one study the deduced hypotheses appeared obtained [32], while in

the other not all results were in accordance with the deduced hypotheses [33]. The studies resulted in fair and poor methodological quality respectively. Evidence synthesis of the WHO HPQ resulted in conflicting evidence for hypotheses testing.

Both studies assessed the criterion validity and specified sensitivity and specificity [32, 33]. In one study [33], the criterion used could be considered an adequate 'gold standard', and the evidence was provided. However, it could only be deduced how missing items were handled. This study resulted in good methodological quality. Area under the ROC curve calculated ranged from 0.63 to 0.69. The other study [32] reported sensitivity and specificity but did not provide a description of how missing items were handled, and it was unclear whether the gold standard was appropriate. The methodological quality of this study is fair. Evidence synthesis of the WHO HPQ resulted in conflicting evidence for criterion validity.

Responsiveness: Because the background of the studies from which the respondents are derived was unclear, and there was lacking information on how missing items were handled, the responsiveness in Kessler et al. [32] results in poor methodological quality. Evidence synthesis resulted in unknown evidence for responsiveness because of the study's poor methodological quality.

There were no methodologically sound studies evaluating the internal consistency, measurement error, cross-cultural validity or responsiveness of the WHO HPQ.

The Stanford Presenteeism Scale (SPS)

The SPS evaluates the impact of health problems on individual performance and productivity by rating six statements on a five-point scale [34, 41].

Reliability: Two studies assessed the internal consistency of the SPS [34, 41]. Both studies performed factor analyses and calculated Cronbach's alpha; 0.80 and 0.82 respectively. Though there was no description of how missing items were handled, it could be deduced. Therefore, the studies result in good methodological quality. Evidence synthesis of the SPS resulted in strong positive evidence for internal consistency (two studies of good methodological quality and positive findings).

Turpin et al. [41] assessed reliability without stating the time interval, and without describing how missing items were handled. The Pearson correlation coefficient was calculated and indicated a strong negative relationship of -0.60 ($P < 0.001$). The methodological quality of the study is fair. Evidence synthesis of the SPS resulted in limited negative evidence for reliability.

Validity: Because Koopman et al. [34] solely assessed discriminative validity for a scale, where none of the relationships showed a strong degree of magnitude, and did not assess whether all items were relevant for the purpose of the instrument. The methodological quality of the study is fair and evidence synthesis of the SPS resulted in limited negative evidence for content validity.

In both studies, (classical) factor analysis was conducted to assess structural validity [34, 41], and indicated that the instrument captured the dimensions intended to assess, providing positive evidence for structural validity. The type of

factor analysis was appropriate in view of the existing information and the way missing items were handled could be deduced. Therefore, the studies are of good methodological quality. Evidence synthesis of the SPS resulted in strong positive evidence for structural validity (two studies of good methodological quality and positive findings).

Because hypotheses were vaguely formulated, the comparator instrument was the long version of the instrument under evaluation, and information on the measurement properties of the comparator were lacking, [34] is fair. The methodological quality for the study by Turpin et al. [41] is also rated fair because solely a number of hypotheses were formulated a priori and the comparator instrument measured another construct (work output versus work impairment). Therefore, one cannot be sure whether the measurement properties of the comparator instrument apply to this study population. Evidence synthesis of the SPS resulted in moderate positive evidence for hypotheses testing (two studies of fair methodological quality and the results were in accordance with the hypotheses).

Both studies assessed criterion validity but did not provide the percentage of missing items [34, 41]. Additionally, Koopman et al. [34] did not provide evidence but only assumed that the criterion used could be considered an adequate gold standard. The continuous scores on both instruments correlated strongly. To assess criterion validity Turpin et al. [41] calculated correlations between SPS metrics and comparator dimensions which were in the expected direction and mostly significantly different from zero. However, Turpin et al. [41] did not make it clear how missing items were handled. Evidence synthesis of the SPS resulted in moderate positive evidence for criterion validity (fair methodological quality and positive findings).

There were no methodologically sound studies evaluating measurement error, cross-cultural validity or responsiveness of the SPS.

The Productivity and Disease Questionnaire (PRODISQ)

In brief, the PRODISQ measures and values productivity costs by assessing the relationship between health and productivity based on open and multiple-choice questions in seven modules [24].

Validity: Koopmanschap et al. [24] assessed whether all items refer to relevant aspects of the constructs to be measured, assessed if all items were relevant to the different study populations, and whether the items were relevant for the application purpose (positive item relevance). Furthermore, the authors assessed whether all items together comprehensively reflected the constructs to be measured (positive item comprehensiveness). Evidence synthesis of the PRODISQ resulted in strong positive evidence for content validity (excellent methodological quality and positive findings).

There were no methodologically sound studies evaluating the internal consistency, reliability, measurement error, structural validity, hypotheses testing, cross-cultural validity or responsiveness of the PRODISQ.

The Health-Related Productivity Questionnaire Diary (HRPQ-D)

The HRPQ-D measures work productivity data related to health-related labor force participation via nine open and multiple-choice questions [35].

Reliability: Kumar et al. [35] investigated the reliability by calculating correlation coefficients (generally non-statistically significant, ranging from -0.449 up to 0.806) between the reported productivity loss data and the symptom scores of repeated measurements within individuals but did not provide evidence as to whether systematic change had occurred. Additionally, several methodological flaws in the design and execution of the study, such as sampling bias, a lack of evidence that the patients were stable and different scoring systems for symptom severity resulted in the study being of fair methodological quality. Evidence synthesis of the HRPQ-D resulted in limited negative evidence for reliability.

Validity: Kumar et al. [35] did not assess whether all items were relevant for the different target populations or whether all items together comprehensively reflected the construct to be measured. The study is therefore of poor methodological quality. Evidence synthesis of the HRPQ-D resulted in unknown evidence for content validity.

A priori hypotheses were vaguely formulated, the expected direction and magnitude of the correlations were poorly described, and no information on the measurement properties of the comparator instrument was given. Generally, the correlations were not statistically significant. The study of Kumar et al. [35] results in poor methodological quality and evidence synthesis of the HRPQ-D resulted in unknown evidence for hypotheses testing.

Kumar et al. [35] did not elaborate on how missing items were handled, and it was unclear whether the criterion used could be considered an adequate gold standard. Therefore, the study results in fair methodological quality. Evidence synthesis of the HRPQ-D resulted in limited negative evidence for criterion validity (fair methodological quality and negative findings).

Responsiveness: Hypotheses were vague, neither magnitudes nor directions of the correlations were stated a priori, and information on the psychometrics of the comparator was lacking. Evaluating the instrument's responsiveness to change yielded positive findings. However, the significant correlation values ranged from -0.161 to 0.422 [35]. Evidence synthesis of the HRPQ-D resulted in limited negative evidence for responsiveness (fair methodological quality and negative findings).

There were no methodologically sound studies evaluating the internal consistency, measurement error, structural validity or cross-cultural validity of the HRPQ-D.

The Health and Labor Questionnaire (HLQ)

The HLQ collects data on the relationship between illness and treatment and work performance. The HLQ is a modular questionnaire with four modules and response options on a four-point scale [20, 25].

Reliability: In the study of Meerding et al. [20], inter-rater reliability was assessed. However, because there was no clear description of how missing items were handled and it was only assumable that participants and test conditions were stable in the period on the construct to be measured, the methodological quality of the study is fair. The agreement on self-reported productivity loss due to health problems showed a poor κ -value of 0.18 [20]. Evidence synthesis of the HLQ resulted in limited negative evidence for reliability.

Validity: Two studies assessed the HLQ's criterion validity [20, 25]. Both studies calculated correlation coefficients (significant correlations ranging from 0.33 to 0.73 [20] and Pearson correlation between 0.41 and 0.56 [25]) However, it was unclear for the study conducted by Van Rooijen, Essink-Bot, Koopmanschap, Bonsel & Rutten [25] whether the criterion used could be considered as a gold standard. Meerding et al. [20] lacked information on how missing items were handled. Therefore, the methodological quality of both studies is fair. Evidence synthesis of the HLQ resulted in moderate negative evidence for criterion validity because of the fair methodological quality of the studies and negative findings (correlations <0.70).

There were no methodologically sound studies evaluating the internal consistency, measurement error, content validity, structural validity, hypotheses testing, cross-cultural validity or responsiveness of the HLQ.

The Quality and Quantity questionnaire (Q&Q)

The Q&Q questionnaire scores two questions about the quantity and quality of the work performed on the last working day on a 10-point scale [20].

Reliability: Meerding, Ijzelenberg, Koopmanschap, Severens & Burdorf [20] did not provide a clear description of how missing items were handled. The stability of participants and test conditions in the period on the construct to be measured could be deduced. The methodological quality of the study is fair. An unweighted kappa (κ -value = 0.18) was calculated to assess the inter-rater reliability [20]. Evidence synthesis of the Q&Q resulted in limited negative evidence for reliability.

Validity: Although the criterion used could be considered an adequate gold standard based on the evidence provided, Meerding, Ijzelenberg, Koopmanschap, Severens & Burdorf [20] lacked a clear description of how missing items were handled. Self-reported productivity measured with the QQ correlated significantly with objective work output, however the strength of the correlation ($r = 0.48$) was poor [20]. Evidence synthesis of the Q&Q resulted in limited negative evidence for criterion validity.

There were no methodologically sound studies evaluating the internal consistency, measurement error, content validity, structural validity, hypotheses testing, cross-cultural validity or responsiveness of the Q&Q.

The Well-Being Assessment for Productivity (WBA-P)

The WBA-P provides an evaluation of job performance loss due to wellbeing-related barriers. The barriers are based on twelve items to be rated from zero to 100 [38].

Validity: Prochaska et al. [38] performed factor analysis to assess the structural validity but did not describe the percentage of missing items. Because it could be deduced how missing items were handled, the methodological quality of the study is good. A hierarchical two-factor model demonstrated good fit ($X^2 = 544.34$) with acceptable internal consistency on the subscales (Cronbach's alpha 0.73 – 0.83) [38]. Evidence synthesis of the WBA-P resulted in limited positive evidence for structural validity because of the paper's good methodological quality.

Prochaska et al. [38] assessed the criterion validity by calculating correlations and multivariate variance analysis. Although no evidence was provided, it was stated that the criterion used could be considered an adequate gold standard. Significant interactions between the WBA-P and the comparator instrument existed [38]. Evidence synthesis of the WBA-P resulted in limited positive evidence for structural validity (good methodological quality and positive findings).

There were no methodologically sound studies evaluating the internal consistency, reliability, measurement error, content validity, hypotheses testing, cross-cultural validity or responsiveness of the WBA-P.

The Health and Work Questionnaire (HWQ)

The HWQ measures workplace productivity and workers' health multidimensionally via 24 questions to be rated on a ten-point scale [40].

Reliability: Shikiar, Halpern, Rentz & Khan [40] conducted factor analysis to check the unidimensionality of the scale. The factor analysis resulted in six factors with eigenvalues greater than 1.0 and accounted for 69% of the variance in the correlation matrix. The sample size was adequate (N=294), and Cronbach's coefficient alpha's were calculated to assess the internal consistency for each subscale separately (ranging from 0.72 to 0.96). The percentage of missing items was not described [40]. The methodological quality of the study is rated fair because it was not clear how missing items were handled. Evidence synthesis of the HWQ resulted in limited positive evidence for internal consistency (fair methodological quality and positive findings: correlations >0.70).

Validity: Shikiar, Halpern, Rentz & Khan [40] assessed the structural validity by conducting a principal components factor analysis with orthogonal rotation confirming that the items on the HWQ are measuring several different dimensions related to productivity and the items on the HWQ do not 'hang together' as well as the items on the individual scales [40]. Because it was not clear how missing items were handled, the methodological quality of the study is fair. Evidence synthesis of the HWQ resulted in limited negative evidence for structural validity.

In assessing the criterion validity Shikiar, Halpern, Rentz & Khan [40] did not make it clear how missing items were handled nor provided evidence, although it

could be deduced, that the criterion used could be considered an adequate gold standard. Correlations with two objective productivity measures ranged from -0.043 to -0.219 and certain subscales did not significantly correlate with overall productivity. Evidence synthesis of the HWQ resulted in limited negative evidence for criterion validity (fair methodological quality and weak correlations).

Responsiveness: Shikiar, Halpern, Rentz & Khan [40] lacked a priori information on the measurement properties of the comparator instrument, the expected direction and magnitude of the correlation. The methodological quality of the study is poor. Evidence synthesis of the HWQ resulted in unknown evidence for responsiveness.

There were no methodologically sound studies evaluating the reliability, measurement error, content validity, hypotheses testing or cross-cultural validity of the HWQ.

The Work and Health Interview (WHI)

In summary, the WHI captures data on absence time from work, reduced performance, and health-related causes based on a six-module, computer-assisted telephone interview [43].

Reliability: Stewart, Ricci, Leotta & Chee [43] studied reliability whereby independent measurements with an appropriate time interval were conducted. Both Pearson's (0.59) and Spearman's (0.63) correlation coefficients were derived between two measures. Evidence was provided that test conditions were similar, it was described how missing items were handled and the sample size (N=66) was adequate [43]. Evidence synthesis of the WHI resulted in limited negative level of evidence for reliability because of the study's good methodological quality and relatively negative findings (correlations <0.70).

Validity: In the study conducted by Stewart, Ricci, Leotta & Chee [43], it could easily be assumed that the criterion used (diary work time measures) could be considered an adequate gold standard. The Pearson's (Spearman's) correlation between the WHI and diary total productivity time measure was 0.50 (0.46) and there were no other important methodological flaws in the design of the study [43]. Evidence synthesis of the WHI resulted in limited negative evidence for reliability (good methodological quality and negative findings).

Responsiveness: Hypotheses were vaguely formulated and limited evidence on the measurement properties of the comparator instrument was available. The methodological quality of the study is fair. The correlations between change scores of time not working at work (continuous scores regarding performance data) were (Pearson's correlation) 0.19 and (Spearman's) 0.33 [43]. Evidence synthesis of the WHI resulted in limited negative evidence for responsiveness.

There were no methodologically sound studies evaluating the internal consistency, measurement error, content validity, structural validity, hypotheses testing or cross-cultural validity of the WHI.

The Valuations of Lost Productivity questionnaire (VOLP)

The VOLP assesses labor input loss due to health via 36 open and multiple-choice questions in six sections [23].

Reliability: Zhang, Bansback, Kopec & Anis [23] assessed reliability over time by means of a test-retest analysis. Kappa statistics were calculated, all being statistically significant and reporting adequate agreement. Due to the time frame (2 weeks), the type of administration (postal questionnaire) and the environment (at home), it could be assumed that the test conditions were similar. The methodological quality of the study is good regarding test-retest reliability. Evidence synthesis of the VOLP resulted in limited positive evidence for reliability.

Validity: In order to assess the constructs of the VOLP, the criterion used could be considered as a reasonable gold standard because evidence regarding the psychometric properties of the comparators is provided in the study by Zhang, Bansback, Kopec & Anis [23]. There is no description on how missing items were handled; however, it could be deduced. Furthermore, Spearman's correlation coefficients were calculated to assess criterion validity. The correlations between the VOLP outcomes and corresponding WPAI outcomes ranged from 0.39 to 0.57. The methodological quality of the paper is good and thus the evidence synthesis of the VOLP resulted in limited negative evidence for reliability (good methodological quality and Spearman correlations <0.7).

There were no methodologically sound studies evaluating the internal consistency, measurement error, content validity, structural validity, hypotheses testing, cross-cultural validity or responsiveness of the VOLP.

Conclusion and discussion

Twenty-five studies on measurement properties of 15 generic self-reported instruments measuring health-related productivity changes have been systematically reviewed, and their methodological quality has been evaluated using the COSMIN-checklist in a best evidence synthesis. The WLQ is the most frequently evaluated instrument. Structural validity and content validity reported a strong and moderate positive level of evidence respectively. For measurement error and cross-cultural validity, no information was available and the internal consistency and criterion validity resulted in conflicting evidence. Reliability, hypotheses testing and responsiveness resulted in limited negative and moderate negative evidence respectively. Due to poor methodological quality, the EWPS, WPAI, and WPSI showed unknown levels of evidence for almost half of the information on measurement properties. For eight questionnaires (AMA-guides, PRODISQ, HLQ, Q&Q, WBA-P, HWQ, WHI, VOLP) at least half of the information on measurement properties per questionnaire was lacking. Four instruments (WLQ, WHO-HPQ, SPS, and PRODISQ) showed

strong or moderate positive levels of evidence for some of the measurement properties.

The main strength of most studies was that they reported detailed information regarding the population characteristics, sampling methods, the setting and the country where the studies were conducted.

There were, however, many limitations. First, the generalizability of the results of the studies on measurement properties was low, mainly because of selective samples, the non-reporting of and the lack of information regarding the handling of missing values, and inadequate sample sizes.

Second, most studies recruited convenience samples, which might not cover the entire target population [44]. Ozminkanski et al. was the only study that consecutively recruited workers with job-related accidents or injuries, which could be a reasonable representation of the workers with lost productivity at the workplace [21].

Third, although Zhang and colleagues examined the measurement properties in two countries (UK and Ireland), no international samples demonstrated the cross-cultural validity of their measures [22, 23]. Most studies were conducted in the United States, which makes it difficult to discern whether the instruments are appropriate for study populations outside of the United States. The results of this review emphasize the need for international studies on measurement properties as well as additional evaluation studies conducted worldwide to examine the cross-cultural appropriateness of these measures to improve generalizability. The Work Role Functioning Questionnaire (WRFQ) measures perceived difficulties in meeting work demands among employees given their physical and emotional problems. The WRFQ addresses work outcomes in an attempt to describe how health affects work role functioning [45-48]. Despite the fact that the WRFQ is to be used as a detection instrument to identify, and not value decreased productivity, it can serve as an excellent example since several studies have translated and adapted the WRFQ to Canadian French [45], Brazilian Portuguese [46], Dutch [47], and Spanish [48]. These studies demonstrate a systematic procedure for translation and cross-cultural adaptation which can serve as excellent examples for future studies attempting to adapt and validate instruments in other cultures.

Fourth, almost half of the reviewed studies reported item and unit nonresponses under 50%, which might indicate selection bias, further hampering the generalizability of the results [44]. Inadequate descriptions of the handling of missing values might suggest non-random missing items, which could bias the results and lead to misinterpretation and misjudgment of the measurement properties of an instrument. Furthermore, if missing values are inappropriately handled, bias in parameter estimates can occur, resulting in lower sample sizes and thus lower statistical power. High percentages of missing values on specific items might even indicate that an item is not relevant for the study population, perhaps pointing to ambiguous formulations and hampering the validity of the instruments [17, 49]. In light of the flaws presented from previous studies, response rates should be accu-

rately reported, including information on the handling of missing items, and if randomness of nonresponse occurred, it should be examined and reported in future studies.

Fifth, based on the results of this systematic review it can be concluded that the information regarding the measurement properties of generic self-reported instruments measuring health-related productivity changes is mostly limited or of poor to fair methodological quality. The results should be treated with caution due to the missing information on the remaining measurement properties. Especially when considering measurement error and cross-cultural validity, wherefore (almost) no information was available.

Sixth, although it is difficult to determine the criterion validity without a real gold standard for health-related productivity change instruments, most studies considered the extent to which scores on the instrument of interest could be adequately reflected to a predetermined comparator. Criterion validity is therefore most frequently evaluated, but only five studies on this measurement property were of good methodological quality. As a consequence, the SPS and the WBA-P yielded moderate and limited positive levels of evidence for criterion validity respectively.

Seventh, it is difficult to determine the responsiveness of the different health-related productivity change instruments because almost all of the retrieved studies were of poor or fair methodological quality regarding responsiveness. Because the instruments are often used as an outcome measure to determine the costs of lost productivity, specific hypotheses regarding expected correlations with other constructs must be formulated a priori when developing a new measure.

Eighth, the internal consistency statistic only gets an interpretable meaning when the interrelatedness among the items is determined as a set of items that together form a reflective model. The internal consistency of an instrument is thus reflected in the quality assessment of structural validity, and vice versa. If the structural validity was not assessed by analyzing the unidimensionality (there is no evidence that the scales are unidimensional), no internal consistency statistic can be properly interpreted. Four studies resulted in good methodological quality on structural validity for the WLQ and SPS, and also in good internal consistency for both instruments. Most of the instruments with unknown levels of evidence due to poor methodological quality regarding internal consistency also lacked information on structural validity.

Ultimately, some general issues on measuring productivity changes should be addressed. First, the concept of productivity loss due to illness is, according to economic theory, based on the concept of a production function where output is a function of capital input, labor input and technology. The focus of most productivity measurement instruments, as has been seen, is on the individual's labor input; measuring the time a person is not at work due to health complaints (absenteeism), or is not productive while at work due to health complaints (presenteeism). However, job and workplace characteristics also play a key role and differ among coun-

tries, which are reflected in the socio-political context in which the study takes place. For example, in some countries that have a workers' compensation system, such as Canada and the United States, there is a differentiation between work and non-work related disability. In other countries, such as the Netherlands, no such differentiation exists. Due to these variations arising from the social-political context, one cannot assume a 'one size fits all' mentality when comparing instrument effectiveness across countries or cultures. Transparency in reporting the key aspects of measurement and validation of health-related productivity would simplify the comparability and usability of the results for occupational and health economic decision making.

Another point to be addressed is that although the COSMIN taxonomy might contribute to a better understanding in the terminology used in validation studies and provides a structured procedure for the evaluation of the methodological quality of the studies on measurement properties, the taxonomy provides a lot of room for interpretation in the checklist items. To minimize interpretability differences between reviewers (CYGN, AER, SMAAE), decisions had to be made on how to score the different items. For example, a problem encountered during the rating of 'criterion validity' was the absence of a gold standard in health-related productivity change instruments. One example of how this problem was dealt with was by assuming the original long version of the shortened instrument being assessed was an adequate comparator, and can thus be seen as a 'gold standard'. Furthermore, since the studies were systematically reviewed on the measurement properties of self-reported instruments which encounter subjective data, it was agreed that objective, registered data could serve as an adequate comparator as well. Predetermined and transparent arguments that the gold standard is 'gold' had to be thoroughly discussed and decided beforehand to assess criterion validity.

Finally, although an agreement was reached that objective data could be seen as a 'gold standard' for collecting lost productivity data it should be in mind that both objective and self-reported instruments have their advantages and disadvantages, which need to be weighed per research question. For example, when using objective insurance data, a particular disadvantage is that the data reflects what has been compensated. What has been compensated does not necessarily reflect the actual time a worker has been unable to work. Productivity changes related to sick leave should therefore always be supplemented by the productivity changes due to decreased work performance; i.e. presenteeism, to avoid underestimations.

Recommendations

Although only cautious advice can be provided on the most appropriate instruments to capture changes in productivity for use in occupational and economic health practice, the WLQ is cautiously recommended at the moment because the

instrument is most frequently evaluated and moderate respectively strong positive evidence was found for content and structural validity respectively. However, negative evidence was found for reliability, hypothesis testing and responsiveness. The WLQ is only used in an English-speaking study population. Using the PRODISQ is cautiously preferred when conducting a study in the Netherlands based on its strong positive evidence for content validity, although evidence for the other measurement properties is lacking. In order to improve the interpretation of the PRODISQ scores, more research regarding the measurement properties (aside from content validity) is needed. The Stanford Presenteeism Scale (SPS) can also be cautiously recommended as it is evaluated in two studies and showed strong positive results for internal consistency and structural validity, and moderate positive results for hypotheses testing and criterion validity. Limited negative evidence however was available for reliability and content validity and information on the other measurement properties was lacking.

Better knowledge and usage of key methodological principles based on quality checklists, such as COSMIN, is recommended to provide high-quality studies evaluating the measurement properties of new and existing instruments in the future. High-quality studies that evaluate and provide strong evidence for the unknown measurement properties, especially cross-cultural validity, are recommended to improve the generalizability and applicability of generic self-reported health-related productivity change instruments. Given the large number of available productivity instruments the development of new instruments is not recommended, but rather improvement of the existing instruments.

List of abbreviations

AMA-guide	American Medical Association-Guides
COSMIN	Consensus-based Standards for the selection of health Measurement Instruments
EWPS	Endicott Work Productivity Scale
HLQ	Health and Labor Questionnaire
HRPQ-D	Health-Related Productivity Questionnaire Diary
HWQ	Health and Work Questionnaire
SPS	Stanford Presenteeism Scale
PRODISQ	PROductivity and Disease Questionnaire
Q&Q	Quality and Quantity questionnaire
VOLP	Valuation of Lost Productivity questionnaire
WBA-P	Well-Being Assessment for Productivity
WHI	Work and Health Interview
WHO HPQ	World Health Organization Health and work Performance Questionnaire
WLQ	Work Limitations Questionnaire
WPAI	Work Productivity and Activity Impairment Instrument
WPSI	Work Productivity Short Inventory

Acknowledgement

The authors would like to thank Casey O'Dell for her help with editing the manuscript for English language errors.

Additional file 1. Search strategy

Medline (PubMed)

Construct Search "productivity" OR "workplace productivity loss" OR "reduced workplace productivity" OR "lost workplace productivity" OR "lost productivity" OR "production capacity" OR "employability" OR "work participation" OR "work performance" OR "work functioning" OR "work limitations" OR "absenteeism" OR "sickness absence" OR "return to work" OR "sick leave" AND "presenteeism" OR "workers' compensation"

Precise search (measurement properties, validation studies) (instrumentation[sh] OR Validation Studies[pt] OR 'reproducibility of results'[MeSH Terms] OR reproducib*[tiab] OR 'psychometrics' OR psychometr*[tiab] OR clinimet*[tiab] OR clinometr*[tiab] OR 'observer variation' OR observer variation[tiab] OR 'discriminant analysis' AND reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR 'internal consistency'[tiab] OR (cronbach[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tiab] OR precision[tiab] OR 'imprecision'[tiab] OR 'precise values'[tiab] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR intratester[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-rater[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR inter-examiner[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tiab] OR ((repli-cab*[tiab] OR repeated[tiab]) AND (measure[tiab] OR measures[tiab] OR findings[tiab] OR result[tiab] OR results[tiab] OR test[tiab] OR tests[tiab])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR 'known group'[tiab] OR factor analysis[tiab] OR factor analyses[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR item discriminant[tiab] OR inter scale correlation*[tiab] OR error[tiab] OR errors[tiab] OR 'individual variability'[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncer-tainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR 'standard error of measurement'[tiab] OR sensitiv*[tiab] OR respon-sive*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR meaningful change[tiab] OR 'ceiling effect'[tiab] OR 'floor effect'[tiab] OR 'item response model'[tiab] OR IRT[tiab] OR Rasch[tiab] OR 'Differential item functioning'[tiab] OR DIF[tiab] OR 'computer adaptive testing'[tiab] OR 'item bank'[tiab] OR 'cross-cultural equivalence'[tiab])

Instrument search 'questionnaires' OR 'questionnaire' OR 'self-report' OR 'self-reports' OR 'generic instruments'

Exclusion filter 'addresses'[Publication Type] OR 'biography'[Publication Type] OR 'case reports'[Publication Type] OR 'comment'[Publication Type] OR 'directory'[Publication Type] OR 'editorial'[Publication Type] OR 'festschrift'[Publication Type] OR 'interview'[Publication Type] OR 'lectures'[Publication Type] OR 'legal cases'[Publication Type] OR 'legislation'[Publication Type] OR 'letter'[Publication Type] OR

	'news'[Publication Type] OR 'newspaper article'[Publication Type] OR 'patient education handout'[Publication Type] OR 'popular works'[Publication Type] OR 'congresses'[Publication Type] OR 'consensus development conference'[Publication Type] OR 'consensus development conference, nih'[Publication Type] OR 'practice guideline'[Publication Type] NOT ('animals'[MeSH Terms] NOT 'humans'[MeSH Terms])
Embase	
Construct Search	'productivity' OR 'workplace productivity loss' OR 'reduced workplace productivity' OR 'lost workplace productivity' OR 'lost productivity' OR 'production capacity' OR 'employability' OR 'work participation' OR 'work performance' OR 'work functioning' OR 'work limitations' OR 'absenteeism' OR 'sickness absence' OR 'return to work' OR 'sick leave' AND 'presenteeism' OR 'workers' compensation'
Precise search (measurement properties, validation studies)	'intermethod comparison'/exp OR 'data collection method'/exp OR 'validation study'/exp OR 'feasibility study'/exp OR 'pilot study'/exp OR 'psychometry'/exp OR 'reproducibility'/exp OR 'reproducibility'/exp OR 'reproducibility'/exp OR 'psychometry':ab,ti OR 'clinimet*':ab,ti OR 'clinomet*':ab,ti OR 'observer variation'/exp OR 'observer variation':ab,ti OR 'discriminant analysis'/exp OR 'validity'/exp OR 'reliab*':ab,ti OR 'valid*':ab,ti OR 'coefficient':ab,ti OR 'internal consistency':ab,ti OR (cronbach*':ab,ti AND ('alpha':ab,ti OR 'alphas':ab,ti)) OR 'item correlation':ab,ti OR 'item correlations':ab,ti OR 'item selection':ab,ti OR 'item selections':ab,ti OR 'item reduction':ab,ti OR 'item reductions':ab,ti OR 'agreement':ab,ti OR 'precision':ab,ti OR 'imprecision':ab,ti OR 'precise values':ab,ti OR 'test-retest':ab,ti OR ('test':ab,ti AND 'retest':ab,ti) OR ('reliab*':ab,ti AND ('test':ab,ti OR 'retest':ab,ti)) OR 'stability':ab,ti OR 'intrater':ab,ti OR 'inter-rater':ab,ti OR 'intrarater':ab,ti OR 'intra-rater':ab,ti OR 'intertester':ab,ti OR 'intratester':ab,ti OR 'intra-tester':ab,ti OR 'inter-observer':ab,ti OR 'inter-observer':ab,ti OR 'intraobserver':ab,ti OR 'intra-observer':ab,ti OR 'intertechnician':ab,ti OR 'inter-technician':ab,ti OR 'intra-technician':ab,ti OR 'intra-observer':ab,ti OR 'interexaminer':ab,ti OR 'inter-examiner':ab,ti OR 'intraexaminer':ab,ti OR 'intra-examiner':ab,ti OR 'interassay':ab,ti OR 'inter-assay':ab,ti OR 'intra-assay':ab,ti OR 'intra-assay':ab,ti OR 'interindividual':ab,ti OR 'inter-individual':ab,ti OR 'intra-individual':ab,ti OR 'intra-individual':ab,ti OR 'interparticipant':ab,ti OR 'inter-participant':ab,ti OR 'intra-participant':ab,ti OR 'intra-participant':ab,ti OR 'kappas':ab,ti OR 'kappa':ab,ti OR 'coefficient of variation':ab,ti OR 'repeatab*':ab,ti OR 'repeatab*':ab,ti OR 'repeated':ab,ti AND ('measure':ab,ti OR 'measures':ab,ti OR 'findings':ab,ti OR 'result':ab,ti OR 'results':ab,ti OR 'test':ab,ti OR 'tests':ab,ti)) OR 'generaliza*':ab,ti OR 'generalisa*':ab,ti OR 'concordance':ab,ti OR ('intra-class':ab,ti AND correlation*':ab,ti) OR 'discriminative':ab,ti OR 'known group':ab,ti OR 'factor analysis':ab,ti OR 'factor analyses':ab,ti OR 'factor structure':ab,ti OR 'factor structures':ab,ti OR 'dimensionality':ab,ti OR 'subscale*':ab,ti OR 'multitrait scaling analysis':ab,ti OR 'multitrait scaling analyses':ab,ti OR 'item discriminant':ab,ti OR 'interscale correlation':ab,ti OR 'interscale correlations':ab,ti OR ('error':ab,ti OR 'errors':ab,ti AND (measure*':ab,ti OR correlat*':ab,ti OR evaluat*':ab,ti OR 'accuracy':ab,ti OR 'accurate':ab,ti OR 'precision':ab,ti OR 'mean':ab,ti)) OR 'individual variability':ab,ti OR 'interval variability':ab,ti OR 'rate variability':ab,ti OR 'variability analysis*':ab,ti OR 'responsiveness':ab,ti OR ('limit':ab,ti AND 'detection':ab,ti) OR 'minimal detectable concentration':ab,ti OR 'interpretab*':ab,ti OR ('small*':ab,ti AND ('real':ab,ti OR 'detectable':ab,ti) AND ('change':ab,ti OR 'difference':ab,ti)) OR 'meaningful change':ab,ti OR 'minimal important change':ab,ti OR 'minimal important difference':ab,ti OR 'minimally important change':ab,ti OR 'minimally important difference':ab,ti OR 'minimal detectable change':ab,ti OR 'minimally detectable change':ab,ti OR 'minimally detectable difference':ab,ti OR 'minimal real change':ab,ti OR 'minimally real change':ab,ti OR 'minimally real difference':ab,ti OR 'minimally real difference':ab,ti OR 'minimally real change':ab,ti

	OR 'minimally real difference':ab,ti OR 'ceiling effect':ab,ti OR 'floor effect':ab,ti OR 'item response model':ab,ti OR 'irt':ab,ti OR 'rasch':ab,ti OR 'differential item functioning':ab,ti OR 'dif':ab,ti OR 'computer adaptive testing':ab,ti OR 'item bank':ab,ti OR 'cross-cultural equivalence':ab,ti
Instrument search	'questionnaires' OR 'questionnaire' OR 'self report' OR 'self reports' OR 'generic instruments'
PsychINFO (EBSCOhost)	
Construct Search	"productivity" OR "workplace productivity loss" OR "reduced workplace productivity" OR "lost workplace productivity" OR "lost productivity" OR "production capacity" OR "employability" OR "work participation" OR "work performance" OR "work functioning" OR "work limitations" OR "absenteeism" OR "sickness absence" OR "return to work" OR "sick leave" AND "presenteeism" OR "workers' compensation"
Precise search (measurement properties, validation studies)	instrumentation OR Validation Studies OR 'reproducibility of results' OR reproducib*OR 'psychometrics' OR psychometr* OR clinimet* OR clinometr* reliab* OR unreliab* OR valid*OR 'internal consistency' OR item OR correlation* OR selection* OR reduction*) OR test-retest OR stability OR interrater OR inter-rater OR intra-rater OR intrarater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR repeatab* replica* OR generaliza*OR generalisa*OR interscale correlation*OR error OR errors OR 'individual variability' OR analysis OR values OR uncertainty measurement OR measuring OR 'standard error of measurement' OR sensitiv* OR 'cross-cultural equivalence'
Instrument search	'questionnaires' OR 'questionnaire' OR 'self report' OR 'self reports' OR 'generic instruments'
EconLit	
Construct Search	"productivity" OR "workplace productivity loss" OR "reduced workplace productivity" OR "lost workplace productivity" OR "lost productivity" OR "production capacity" OR "employability" OR "work participation" OR "work performance" OR "work functioning" OR "work limitations" OR "absenteeism" OR "sickness absence" OR "return to work" OR "sick leave" AND "presenteeism" OR "workers' compensation"
Precise search (measurement properties, validation studies)	instrumentation OR Validation Studies OR 'reproducibility of results' OR reproducib*OR 'psychometrics' OR psychometr* OR clinimet* OR clinometr*AND reliab* OR unreliab* OR valid*OR 'internal consistency' OR item OR correlation* OR selection* OR reduction*) OR test-retest OR stability OR integrate OR inter-rater OR intra rater OR intrarater OR interested OR inter-tester OR indra tester OR indra-tester OR inter Observer OR inter-observer OR indra Observer OR indra-observer OR repeatab*
Instrument search	'questionnaires' OR 'questionnaire' OR 'self report' OR 'self reports' OR 'generic instruments'

Additional file 2. Detailed description of the self-reported instruments measuring health-related productivity changes

Instrument	Construct / Content	# questions / items	Rating	Item example	Recall period	Discipline
Work Limitations Questionnaire (WLQ) (18, 20, 26, 27, 35, 36, 41, 43)	Degree to which health problems interfere with specific aspects of job performance and its productivity impact	16 or 25 4 dimensions: physical demands, time management, output demands, and mental/interpersonal demands	5-point scale from none of the time (0%) to all of the time (100%)	"In the past two weeks, how much of the time did your physical health or emotional problems make it difficult for you to stick to a routine or schedule?"	2 to 4 weeks	Medical
Endicott Work Productivity Scale (EWPS) (18, 28)	Wide variety of mental and medical disorders working in a wide variety of job settings to assess the degree to which this affects the work functioning of a subject	25	5-point scale of how often the behaviour or feeling, or attitude has been manifested during the past week: 0 (best possible score) to 100 (worst possible score)	"During the past week, how frequently did you find you have forgotten to call someone?"	7 days	Medical
Work Productivity and Activity Impairment Instrument (WPAI) (18, 21, 38)	The effect of general health and symptom severity on work productivity and regular activities	6	5-point scale of how severity of symptoms interfere with performance: 0 ('very severe') to 5 ('did not have symptom')	"Are you currently employed (working for pay)?"	7 days	Economics
Work Productivity Short Inventory (WPSI) (18, 20, 30)	The prevalence of medical problems that may influence work productivity and the financial implications thereof	22	Open questions	"During the past year, did you experience an anxiety disorder, such as Generalized Anxiety Disorder, Panic Disorder, Phobias, Obsessive Compulsive Disorder, or Post-traumatic Stress Disorder?"	2 weeks, 3 months, 1 year	Social sciences
American Medical Association-Guides (AMA-guide) (29)	The loss of function and determination of compensation and ability to work after injury or illness	15	10-point scale how pain now affects functioning in everyday activities: 0 ('work normally') to 10 ('unable to work at all')	"Does your pain interfere with your normal work inside and outside the home?"	Now	Medical

Instrument	Construct / Content	# questions / items	Rating	Item example	Recall period	Discipline
World Health Organization Health and work Performance Questionnaire (WHO HPQ) (31, 32)	Indirect workplace costs of illness by assessing absenteeism, presenteeism, and critical incidents	3 categories: Health (31 items) Work (16 items) Demographics (8 items)	Open questions	"Now please think of your work experiences over the past 4 weeks (28 days). In the spaces provided below, write the number of days you spent in each of the following work situations. In the past 4 weeks (28 days), how many days did you miss an entire work day because of problems with your physical or mental health? Number of days (00-28)."	Four week period or the past 7 days	Social sciences (occupational environmental medicine)
Stanford Presenteeism Scale (SPS) (33, 40)	The impact of health problems on individual performance and productivity	6 statements reflecting cognitive, emotional and behavioural aspects	Five-point scale from 1 (if you strongly disagree with the statement) to 5 (if you strongly agree with the statement)	"Statement: 'because of my health problem, the stresses of my job were much harder to handle.'"	30 days	Social sciences (occupational environmental medicine)
The PROductivity and DISease Questionnaire (PRODISQ) (23)	Productivity costs by assessing the relationship between health and productivity	7 modules 64 questions: General information (N=4), Profession, working situation and income (N=16), Absenteeism (N=4), Compensation mechanisms (N=5), Presenteeism (N=3), Organizational level (N=17), Administrative and management costs (N=15)	Open questions Multiple-choice	"On the scale below, please circle how much work you've done during your last day in your normal working hours compared to a normal working day '0' (nothing) to '10' (as much as normal)."	3 months	Economics
Health-Related Productivity Questionnaire Diary (HRPQ-D) (34)	Work productivity data within clinical trials and survey data collections: data relating to health-related labor force participation	9	Open questions Multiple-choice	"Hours of work outside the home missed because of 'the health problem': record the number of hours you work you missed today because of 'the health problem'. If none, put a '0'	Daily	Social sciences (occupational environmental medicine)

Instrument	Construct / Content	# questions / items	Rating	Item example	Recall period	Discipline
Health and Labor Questionnaire (HLQ) (19, 24)	Relationship between illness and treatment and work performance: estimate the production losses (costs) and the ability to use work as an indicator of health status in the social domain	4 modules: Absence from paid work, Reduced productivity at paid work, Unpaid production, and Impediments to paid and unpaid labor	4-point scale from (almost) never to (almost) always	"I did go to work but as a result of health problems. I had a problem concentrating."	2 weeks	Economics
Quality & Quantity Questionnaire (Q&Q) (19)	Quantity and quality of the work performed on the last working day	2	10-point scale from 0 (practically nothing and very poor quality) to 10 (normal quantity and normal quality)	"Could you indicate how much work you actually performed today during regular hours as compared to normal on the scale below?"	Last working day	Economics
Well-Being Assessment for Productivity (WBA-P) (37)	Informative evaluation of job performance loss due to well-being related barriers	12 common barriers to productivity	Individual's ability to perform his or her best at work (not at all, some, or a lot) Work-related and personal barriers (scoring from 0 'not at all' to 100 'a lot for all 11 reasons')	"How often have you had trouble at work concentrating or doing you best because of ...? (each of the 11 items)"	4 weeks	Social sciences
Health and Work Questionnaire (HWQ) (39)	Workplace productivity and worker health	24	Multi-part questions on a 10 point-scale	"How satisfied were you overall with your job this week?"	7 days	Social sciences
Work and Health Interview (WHI) (42)	Captures data on absence time from work and related causes and on health-related reduced performance while at work and the health-related causes	Six module interview (first 3: verbal informed consent, detailed data on employment status and usual work time, data on episodes of 21 health problems; then: job visualization module, questions about health problems reported as the	Computer-assisted telephone interview (open questions and categorical response options (percentages))	Primes respondents to report on health-related lost productive time on days at work; asks detailed questions about usual tasks and activities performed at work and time allotted to each	2 weeks	Economics

cause of missed work or reduced productivity, and data on demographics and annual salary)

Instrument	Construct / Content	# questions / items	Rating	Item example	Recall period	Discipline
Valuation of Lost Productivity questionnaire (VOLP) (22)	Assess labor input loss due to health (any physical, mental, or emotional problems or symptoms)	36 questions in 6 sections: employment status, job characteristics, absenteeism, work performance, unpaid work, and working environment	Open questions and multiple-choice	"Which of the following describes your current work situation? 1. Working full time as an employee 2. Working part time as an employee 3. Self-employed"	7 days	Social sciences

References

1. Lambeek LC, van Tulder MW, Swinkels IC, Koppes LL, Anema JR, van Mechelen W. The trend in total cost of back pain in The Netherlands in the period 2002 to 2007. *Spine (Phila Pa 1976)* 2011;36:1050-8.
2. Stock S, Redaelli M, Luengen M, Wendland G, Civello D, Lauterbach KW. Asthma: prevalence and cost of illness. *The European respiratory journal* 2005;25:47-53.
3. van Tulder MW, Koes BW, Bouter LM. A cost-of-illness study of back pain in The Netherlands. *Pain* 1995;62:233-40.
4. Goetzel RZ, Hawkins K, Ozminkowski RJ, Wang S. The health and productivity cost burden of the "top 10" physical and mental health conditions affecting six large U.S. employers in 1999. *J Occup Environ Med* 2003;45:5-14.
5. Drummond M, McGuire A. *Economic Evaluation in Health Care. Merging theory with practice.* New York: Oxford University Press; 2001.
6. Schultz AB, Chen CY, Edington DW. The cost and impact of health conditions on presenteeism to employers: a review of the literature. *Pharmacoeconomics* 2009;27:365-78.
7. Smit F, Willemsse G, Koopmanschap M, Onrust S, Cuijpers P, Beekman A. Cost effectiveness of preventing depression in primary care patients: randomised trial. *The British journal of psychiatry : the journal of mental science* 2006;188:330-6.
8. Brouwer WB, Koopmanschap MA, Rutten FF. Productivity losses without absence: measurement validation and empirical evidence. *Health Policy* 1999;48:13-27.
9. Hoeijenbos M, Bekkering T, Lamers L, Hendriks E, van Tulder M, Koopmanschap M. Cost effectiveness of an active implementation strategy for the Dutch physiotherapy guideline for low back pain. *Health policy* 2005;75:85-98.
10. Lotters F, Meerding WJ, Burdorf A. Reduced productivity after sickness absence due to musculoskeletal disorders and its relation to health outcomes. *Scand J Work Environ Health* 2005;31:367-74.
11. Pauly MV, Nicholson S, Polsky D, Berger ML, Sharda C. Valuing reductions in on-the-job illness: 'presenteeism' from managerial and economic perspectives. *Health economics* 2008;17:469-85.
12. Lofland JH, Pizzi L, Frick KD. A review of health-related workplace productivity loss instruments. *Pharmacoeconomics* 2004;22:165-84.
13. Prasad M, Wahlqvist P, Shikier R, Shih YC. A review of self-report instruments measuring health-related work productivity: a patient-reported outcomes perspective. *Pharmacoeconomics* 2004;22:225-44.
14. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2011;
15. Terwee CB, Jansma EP, Riphagen, II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009;18:1115-23.
16. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737-45.
17. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539-49.
18. Furlan AD, Pennick V, Bombardier C, van Tulder M, Editorial Board CBRG. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine (Phila Pa 1976)* 2009;34:1929-41.

CHAPTER 2

19. Erickson SR, Guthrie S, VanEtten-Lee M, Himle J, Hoffman J, Santos SF, Janeck AS, Zivin K, Abelson JL. Severity of anxiety and work-related outcomes of patients with anxiety disorders. *Depression and Anxiety* 2009;26:1165-71.
20. Meerding WJ, W IJ, Koopmanschap MA, Severens JL, Burdorf A. Health problems lead to considerable productivity loss at work among workers with high physical load jobs. *J Clin Epidemiol* 2005;58:517-23.
21. Ozminkowski RJ, Goetzel RZ, Chang S, Long S. The application of two health and productivity instruments at a large employer. *Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine* 2004;46:635-48.
22. Zhang W, Bansback N, Boonen A, Young A, Singh A, Anis AH. Validity of the work productivity and activity impairment questionnaire--general health version in patients with rheumatoid arthritis. *Arthritis Res Ther* 2010;12:R177.
23. Zhang W, Bansback N, Kopec J, Anis AH. Measuring time input loss among patients with rheumatoid arthritis: validity and reliability of the Valuation of Lost Productivity questionnaire. *J Occup Environ Med* 2011;53:530-6.
24. Koopmanschap MA. PRODISQ: a modular questionnaire on productivity and disease for economic evaluation studies. *Expert Rev Pharmacoecon Outcomes Res* 2005;5:23-8.
25. van Roijen L, Essink-Bot ML, Koopmanschap MA, Bonsel G, Rutten FF. Labor and health status in economic evaluation of health care. *The Health and Labor Questionnaire. International journal of technology assessment in health care* 1996;12:405-15.
26. Beaton DE, Kennedy CA. Beyond return to work: Testing a measure of at-work disability in workers with musculoskeletal pain. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation* 2005;14:1869-79.
27. Tang K, Pitts S, Solway S, Beaton D. Comparison of the psychometric properties of four at-work disability measures in workers with shoulder or elbow disorders. *J Occup Rehabil* 2009;19:142-54.
28. Roy JS, MacDermid JC, Amick BC, Shannon HS, McMurtry R, Roth JH, Grewal R, Tang K, Beaton D. Validity and responsiveness of presenteeism scales in chronic work-related upper-extremity disorders. *Phys Ther* 2011;91:254-66.
29. Endicott J, Nee J. Endicott Work Productivity Scale (EWPS): a new measure to assess treatment effects. *Psychopharmacol Bull* 1997;33:13-6.
30. Forst L, Friedman L, Chukwu A. Reliability of the AMA Guides to the Evaluation of Permanent Impairment. *J Occup Environ Med* 2010;52:1201-3.
31. Goetzel RZ, Ozminkowski RJ, Long SR. Development and reliability analysis of the Work Productivity Short Inventory (WPSI) instrument measuring employee health and productivity. *Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine* 2003;45:743-62.
32. Kessler RC, Ames M, Hymel PA, Loeppke R, McKenas DK, Richling DE, Stang PE, Ustun TB. Using the World Health Organization Health and Work Performance Questionnaire (HPQ) to evaluate the indirect workplace costs of illness. *J Occup Environ Med* 2004;46:S23-37.
33. Kessler RC, Barber C, Beck A, Berglund P, Cleary PD, McKenas D, Pronk N, Simon G, Stang P, Ustun TB, Wang P. The World Health Organization Health and Work Performance Questionnaire (HPQ). *Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine* 2003;45:156-74.
34. Koopman C, Pelletier KR, Murray JF, Sharda CE, Berger ML, Turpin RS, Hackleman P, Gibson P, Holmes DM, Bendel T. Stanford presenteeism scale: health status and employee productivity. *Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine* 2002;44:14-20.
35. Kumar RN, Hass SL, Li JZ, Nickens DJ, Daenzer CL, Wathen LK. Validation of the Health-Related Productivity Questionnaire Diary (HRPQ-D) on a sample of patients with infectious mononucleosis: results from a phase 1 multicenter clinical trial. *J Occup Environ Med* 2003;45:899-907.
36. Lerner D, Amick BC, Rogers WH, Malspeis S, Bungay K, Cynn D. The Work Limitations Questionnaire. *Medical care* 2001;39:72-85.

MEASURING HEALTH-RELATED PRODUCTIVITY CHANGES

37. Lerner D, Reed JI, Massarotti E, Wester LM, Burke TA. The Work Limitations Questionnaire's validity and reliability among patients with osteoarthritis. *Journal of clinical epidemiology* 2002;55:197-208.
38. Prochaska JO, Evers KE, Johnson JL, Castle PH, Prochaska JM, Sears LE, Rula EY, Pope JE. The well-being assessment for productivity: a well-being approach to presenteeism. *J Occup Environ Med* 2011;53:735-42.
39. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. *Pharmacoeconomics* 1993;4:353-65.
40. Shiklar R, Halpern MT, Rentz AM, Khan ZM. Development of the Health and Work Questionnaire (HWQ): An instrument for assessing workplace productivity in relation to worker health. *Work: Journal of Prevention, Assessment & Rehabilitation* 2004;22:219-29.
41. Turpin RS, Ozminkowski RJ, Sharda CE, Collins JJ, Berger ML, Billotti GM, Baase CM, Olson MJ, Nicholson S. Reliability and validity of the Stanford Presenteeism Scale. *J Occup Environ Med* 2004;46:1123-33.
42. Walker N, Michaud K, Wolfe F. Work limitations among working persons with rheumatoid arthritis: results, reliability, and validity of the work limitations questionnaire in 836 patients. *The Journal of rheumatology* 2005;32:1006-12.
43. Stewart WF, Ricci JA, Leotta C, Chee E. Validation of the work and health interview. *Pharmacoeconomics* 2004;22:1127-40.
44. Carley S, Lecky F. Statistical consideration for research. *Emergency medicine journal* 2003;20:258-62.
45. Durand MJ, Vachon B, Hong QN, Imbeau D, Amick BC, 3rd, Loisel P. The cross-cultural adaptation of the Work Role Functioning Questionnaire in Canadian French. *Int J Rehabil Res* 2004;27:261-8.
46. Gallasch CH, Alexandre NM, Amick B. Cross-cultural adaptation, reliability, and validity of the work role functioning questionnaire to Brazilian Portuguese. *J Occup Rehabil* 2007;17:701-11.
47. Abma FI, Amick BC, Brouwer S, van der Klink JJ, Bultmann U. The cross-cultural adaptation of the work role functioning questionnaire to Dutch. *Work* 2012;43:203-10.
48. Ramada JM, Serra C, Amick BC, Castano JR, Delclos GL. Cross-cultural adaptation of the Work Role Functioning Questionnaire to Spanish spoken in Spain. *J Occup Rehabil* 2013;23:566-75.
49. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychological methods* 2002;7:147-77.

CHAPTER

3

Patient self-reported and administrative data used for cost estimations of healthcare resource use: A systematic review of methodological reporting quality and validation evidence

Cindy Noben
Angelique de Rijk
Frans Nijhuis
Jan Kottner
Silvia Evers

Abstract

Objective: To assess the exchangeability of self-reported and administrative healthcare resource use measurements for cost estimation.

Study Design: A systematic review (NHS EED & MEDLINE) was conducted. Reviewers working in duplicate evaluated the methodological reporting quality of studies comparing the validation evidence of healthcare resource use measurement instruments. An appraisal tool called MERQ (Methodological Reporting Quality) was developed by merging aspects from the Guidelines for Reporting Reliability and Agreement Studies (GRASS) and The Standards for Reporting Diagnostic Accuracy (STARD).

Results: Out of 173 studies, 35 full-text articles were assessed for eligibility, of which 16 articles were included in this study. In seven articles, more than 75% of the reporting criteria assessed by the quality appraisal tool were considered 'good'. The majority of studies scored at least 'fair' on most of the reporting quality criteria. In the end, six studies scored 'good' on the minimal criteria for reporting. Varying levels of agreement among the different data sources were found, with correlations ranging from 0.14 up to 0.93 and with occurrences of both random and systematic errors.

Conclusion: The validation evidence of the small number of studies with adequate methodological reporting quality cautiously supports the exchangeability of both the self-reported and administrative resource use measurement methods.

Keywords: healthcare, utilization, self-reports, data collection, validation studies, systematic review.

First Revision: Journal of Clinical Epidemiology

Introduction

Essential to the economic evaluation of healthcare treatments is a robust method for collecting data on healthcare resource use, which can be used to calculate the related costs. Healthcare resource quantities can be defined in different ways. The cost types for potential inclusion in an economic evaluation are commonly classified into healthcare costs, patient and family costs, and costs in other sectors (absenteeism and presenteeism) [1, 2]. Universally recognized methods of collecting resource use data are lacking, however, and a wide variety of methods can be used. Resource use data is most often measured through patient self-reports (e.g. cost questionnaires, diaries, interviews) or through administrative records (e.g. administrative systems of care providers or insurance companies).

Both methods have advantages and disadvantages. One should, for instance, be cautious when calculating costs from patient self-reports, especially because patients may not know or be able to recall precisely their length of stay and diagnoses, two of the primary cost drivers [3]. Furthermore, patient self-reports are more burdensome for patients, and the different timing of data collection might influence accuracy. Although there is no common reference standard for measuring resource use, in general, administrative data collection methods are assumed to be the most objective and therefore the most accurate source of information. However, as healthcare provision has become more integrated, economic evaluations increasingly need to include information from many different care providers. Accordingly, it is becoming more difficult and expensive to review the records from the many providers involved in administering treatments to patients. Furthermore, given prolonged survival rates and the ageing population, it is expensive to obtain longitudinal treatment information by accessing medical records at multiple points in time. Even when administrative data are available and accurate, limits on benefit coverage might require the collection of healthcare utilization data from other sources (i.e. alternative healthcare and out-of-pocket payments) to calculate the economic costs. The major advantages and disadvantages of patient self-reported versus administrative measurements are listed in Table 1.

Table 1. Comparing the advantages and disadvantages of patient self-reported versus administrative measurements of resource use

	Advantages	Disadvantages
Self-reports	Low costs	Selective non-response
	Possibility to include all societal costs (formal and informal, reimbursed and non-reimbursed)	Lower compliance in completing over a longer period of time
	Higher visibility of additional non-specific costs (not directly related to the disease or intervention)	Burdensome for patients Tendency to record more socially acceptable answers (respondent bias, recall bias)
Administrative data	No burden for patients	Limited feasibility (privacy, time lag)
	Data already registered	Inconsistent coding/reporting
	Detectable and detailed information on type and volume of care delivered	Cooperation of many different databases necessary
		Costs levied to access the data
		No visibility of additional resource use not directly related to the healthcare institution and/or insurance company (no routine records for patient and family resource use, additional expenditures, input of informal caregivers, presenteeism, etc.) Impossible to include unrelated costs, informal costs and out-of-pocket expenses.

Considerable research effort has been concentrated on the appropriate measurement of outcomes and the policy implications of economic evaluation, but the methods for collecting the resource use data have been relatively neglected [4], and no standard methodology for estimating healthcare resource use has been developed. Consequently, the choice for a collection method for resource use data is often a matter of practicality instead of a decision based on methodological considerations. Earlier review studies primarily focused on the accuracy of patient recall and the impact of patient self-reported methods on the accuracy of estimates [3, 5, 6], but not on the methodological reporting quality of these studies. Due to different definitions and the lack of standard statistics regarding the validation evidence (e.g. agreement, reliability, validity and accuracy) as well as inconsistent methodological reporting quality, caution is warranted when interpreting the study findings.

Similar to all other areas of empirical research, high methodological reporting quality is a prerequisite for interpreting the validation evidence provided. The aim of this systematic review is therefore to assess, based on their Methodological Reporting Quality (MeRQ), whether or not the validation evidence of studies on

healthcare resource use measurements that are intended for cost estimation supports the exchangeability of both self-reported and administrative methods of data collection. The research questions are:

1. *What are the characteristics of studies that compare self-reported with administrative methods of measuring healthcare resource use?*
2. *What is the methodological reporting quality of the studies from which this evidence is derived?*
3. *What validation evidence has been reported when comparing self-reported with administrative methods of collecting healthcare resource use data that is intended for cost estimation?*
4. *Given adequate methodological reporting quality, does the validation evidence support the exchangeability of both self-reported and administrative measurement methods for resource use?*

Methods

This systematic review was planned, conducted, and reported in adherence to the PRISMA reporting guidelines [7].

Study eligibility

Original research in any language that had as a stated purpose the comparison of resource use data gathered using patient self-reports with data gathered using administrative methods for application in costing studies is included in this review. Identification of healthcare resource use should serve as an item for costing and application in trial-based economic evaluations. Therefore, studies that solely modeled or described patterns of healthcare use or studies that examined the impact of various patients' characteristics on congruency were excluded from the analysis. Only studies that contained validation arguments in the interpretation of each study were eligible. Minimal requirements for validation arguments included: identification of healthcare resource use as an item for costing; identification of collection methods for resource use data; and identification of validation evidence (i.e. agreement, reliability and validity). Due to the lack of consistent terminology, some uncertainty existed as to the appropriate application of these specific terms. In cases of conflict, consensus was reached through discussion among the reviewers.

Patient self-reports were considered if they were made solely by the patient; articles using proxies for self-reports were therefore excluded. Any type of direct healthcare resource use (e.g. a visit to a clinical provider, medication usage) and any type of data collection instrument or method (e.g. diary, interview, questionnaire, GP record, insurance database) were considered.

Study identification

A search in NHS EED, by means of the Cochrane Library or the Center for Reviews and Dissemination, along with a supplementary search in MEDLINE (PubMed) was conducted, which is generally an appropriate strategy for identifying all published economic evaluation studies related to healthcare that are relevant [8, 9]. Details of the search strategy are provided in Appendix 2. Both text words and MeSH terms were combined with AND and OR for 'healthcare', 'utilization', 'patient self-report', 'administrative data', 'cost analysis', 'validity', 'accuracy', 'reliability', 'agreement'. Reference and citation lists of included studies were hand searched to identify additional studies. International experts were consulted via the *International Health Economics Discussion List* in order to identify a wider range of relevant studies. An archive of collection methods for healthcare utilization data called DIRUM (Database of Instruments for Resource Use Measurement) [4] was used to test the sensitivity of the search. Titles and abstracts were screened for inclusion, and full texts were screened once the articles were judged eligible for inclusion or uncertain. In cases of conflict, consensus was reached through discussion among the reviewers.

Instrument for methodological reporting quality

A new tool for appraising the methodological reporting quality (MeRQ) for healthcare resource use measurement was developed on the basis of two guidelines for reporting: Guidelines for Reporting Reliability and Agreement Studies (GRRAS) [10] and The Standards for Reporting Diagnostic Accuracy (STARD) [11]. Although STARD focuses on improving the reporting quality of studies on diagnostic accuracy, some items in the checklist refer to a 'reference test', or 'golden standard', which is less relevant for research on healthcare resource use. Diagnostic accuracy studies use a recognized reference standard, whereas reference standards regarding data collection methods for healthcare resource use are not available. Nonetheless, overlap was found in most reporting domains. GRRAS complimented the STARD by emphasizing items specific to reliability studies. An operational adaptation of both guidelines into an appraisal tool for evaluating the methodological reporting quality (MeRQ) can be found in Appendix 1. Adaptations were made by focusing on the methodological content and by deleting the directions and layout positions.

MeRQ was rated using the scoring criteria: good (explicit reporting in the study); fair (reporting makes it possible to deduce what is expected); poor (not explicitly reported and difficult to deduce what is expected), not reported (no information provided or to be deduced, also unregistered, unreported (for example: no information on the measurement procedure)), not addressed (not designating the measurement, also undirected (for example: information on the measurement procedures, though not clear whether they relate to the self-reports or the administrative data)), or not applicable. For each study, the quality of reporting on following criteria was assessed: (i) explicit identification of the study as an evaluation of

the reliability, agreement or validity (accuracy) of a self-reported healthcare resource use measurement instrument compared with healthcare resource use measured using administrative data; (ii) validation arguments and strategies for interpreting validation evidence; (iii) measurement methods (eligibility criteria) with a critical review of evidence relevant for the study; (iv) rationale for the relationship between the self-reports and the administrative data; (v) methods or measurement procedures for the self-reports; (vi) methods or measurement procedures for the administrative data; (vii) time interval between self-reports and administrative data collection; (viii) population characteristics; (ix) statistical methods for validation estimates; (x) validation estimates; (xi) methods for statistical uncertainty; (xii) concluding validation argument and/or interpretation of validation evidence. All authors contributed to the rating of the methodological reporting quality of the studies. In cases of conflicting scores, consensus was reached through discussion.

Validation evidence

There are different concepts in the literature regarding the assessment of measurement validity (i.e. reliability, agreement, validity and accuracy) [12]. In this systematic review, validation evidence is considered an umbrella term whereby the terms reliability, agreement and validity (and accuracy as a sub-concept of validity) are used to provide information about the quality of the measurements. Reliability is defined here as the ratio of variability between the subjects to the total variability of all measurements in the sample. Reliability therefore reflects the ability of a measurement to differentiate between subjects or objects (i.e. discriminative ability). Statistical approaches include intraclass correlation and kappa calculations. Agreement is defined as the degree to which scores are identical and is often expressed in percentages (i.e. proportions of agreement, ranges, limits of agreement). Diagnostic accuracy refers to the degree of agreement between the information from the test being evaluated (index test) and reference standard, and can be expressed in many ways, i.e. sensitivity, specificity, likelihood ratios, odds ratio, area under a receiver operating curve (ROC). For costing measures, diagnostic accuracy is undefined because no reference standard is considered the best available method for the collection of data on resource use. However, as mentioned in the introduction, administrative data collection methods are assumed to be the most accurate source of information and therefore often considered the gold standard. Despite this anomaly, information on accuracy is provided in this study and considered informative to support the validity of the cost estimates. Accuracy is seen as a sub-concept of validity. In this framework, diagnostic accuracy is undefined because in costing measures, no reference standard is considered the best available method for the collection of data on resource use. Though this information is considered informative to support the validity of the cost estimates, accuracy, if at all considered, is in this study seen as a sub-concept of validity. Complete and accurate re-

porting makes it possible to detect the potential for bias in the study (internal validity) and to evaluate the generalizability and applicability of the results (external validity).

Best evidence

Studies with 'good' reporting on four 'minimal criteria for reporting' of MeRQ were compared in detail. These minimal criteria include: clear description of the measurement methods (eligibility criteria) with a critical review of evidence relevant for the study; rationale for the relationship between the self-report and the administrative data; methods or measurement procedures for self-reports; and methods or measurement procedures for the administrative data. A 'good' score on reporting for these criteria was required to discuss whether or not the validation evidence of the studies supported the exchangeability of self-reported and administrative measurements of resource use.

Data extraction

Data was extracted independently. Basic information was extracted on the year of publication and country of origin, study design (e.g. sample of enrollees, case-control study, RCT), description of the population (e.g. workers, patients/disease group) and the sample size, the type of healthcare resource use measured (e.g. medication, GP visits, hospital admissions), the method of self-report (e.g. questionnaire, diary), the type of administrative measure (e.g. hospital record, insurance record), and the data collection period. See Table 2. The MeRQ score is reported per study by presenting the score on each of the reporting criteria (Table 3). The scoring was abstracted independently and in duplicate for all studies and all criteria, and conflicts were resolved by consensus. Validation evidence, as presented in Table 4, included the main statistical analysis, the validation estimates and the statistical uncertainty. In the end, studies that scored 'good' on four 'minimal criteria for reporting' of MeRQ were compared in detail.

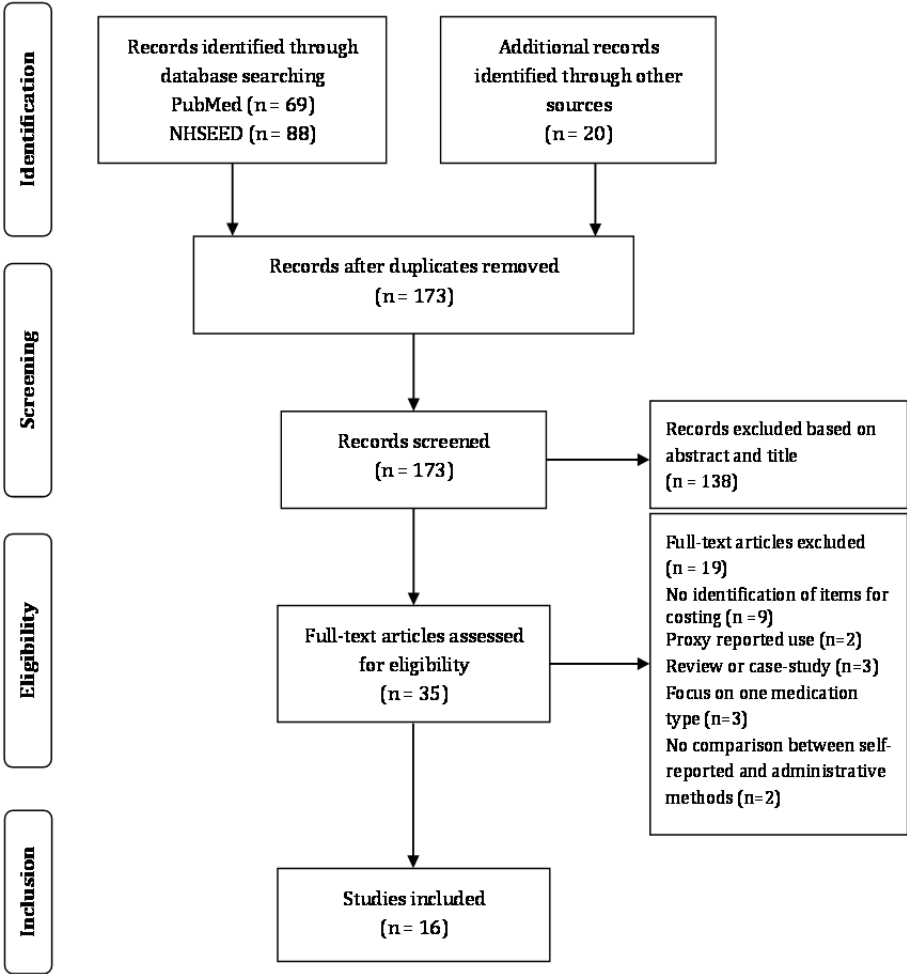
Results

Study selection

Our search in MEDLINE (PubMed) resulted in 69 hits, and 88 hits were found in the NHS Economic Evaluation Database, which is included in The Cochrane Library. After consulting experts, 11 articles were added. Citation tracking and hand searching resulted in 9 additional hits. Four duplicates were removed, and the 173 remaining records were screened by title and abstract. At this stage, 138 records

were excluded. The main reasons for exclusion were that the studies did not aim to measure healthcare resource use for the purpose of costing analysis, did not provide information on the estimation of healthcare costs, and did not address the economic significance or the impact on the costs of care but solely described patterns of healthcare use or provided information on how population characteristics influenced the agreement or disagreement between both measurement methods. Thirty-five full-text articles were assessed for eligibility, whereof another 19 full-text articles were excluded. The most common reasons for exclusion were that the studies reported proxy-recorded resource use; the main aim was not the comparison of healthcare resource use measured via self-reports with use measured via administrative methods; the studies did not measure healthcare resource use as items of costing or did not address the economic impact on costs of care; the studies were a case-study or review; or the resource use was solely focused on the usage of one specific type of medication. In the end, 16 articles were included in this comprehensive literature review [13-28]. A flow chart describes the process of inclusion (see Figure 1).

Figure 1. Flowchart of inclusion



Characteristics of the studies

Table 2 presents an overview of the characteristics of the selected studies (N=16). The majority of studies examined at least one of the following types of resource use: specialist care [13-19, 21, 22, 25, 26, 28], primary care [13, 17, 22-24], medication use [15, 24, 25, 27, 28] or hospitalization [13, 16, 17, 19-21, 26, 28]. Most studies collected information on multiple types of resource use [13, 15-17, 19, 21, 22, 24-26]. Eleven studies were conducted in Europe [13, 14, 16-18, 21-23, 26-28], two in Canada [15, 19], one in the USA [20], one in Australia [25] and one in New Zealand [24]. The population characteristics varied from patients with mental health com-

plaints [13, 21, 22, 28] to musculoskeletal disorders [14, 18, 24], chronic disorders [15, 16, 19, 20], primary care attenders [23], acute ward patients [26] and cancer patients [17, 25, 27]. In most studies, a randomized controlled trial was conducted [14, 16-18, 22, 24-26, 28], of which two were conducted in a multi-center setting [13, 27]. In four studies, the study design was based on a sample of enrollees from another study [21], program [20] or healthcare service [15, 23]. One study used a population-based cohort [19].

In eight studies, healthcare resource use was measured with a questionnaire [15, 18, 22-24, 26-28], in three studies with an interview [13, 19, 21], in another three studies with a diary [14, 25, 27], in two studies with a telephone survey [17, 20] and one with a booklet [16]. The studies that measured healthcare resource use based on administrative records used data from medical records [13], including care provider registrations [15, 16, 21], record forms [17, 18, 22-24, 27] and provider reports [26]. Six studies used health insurance data [14, 15, 19, 20, 25, 28].

In most studies, the data collection was retrospective [13, 17-24, 26, 28] with the largest recall period of twelve months [13, 18, 19]. In three studies that used diaries for self-reports, the data was collected prospectively [14, 16, 25]. One study collected the data on a daily basis [15]. Another study collected the data both prospectively (via a diary) and retrospectively (via questionnaires) [27].

Table 2. Characteristics of studies comparing measurement methods for healthcare resource use

Reference & year	Country	Study design	Population (sample size)	Healthcare resource use	Method of self-report	Administrative record	Data collection period
Byford (2007) [13]	England & Scotland	Multi-center RCT	Adults with recurrent episodes of deliberate self-harm (N=272)	Inpatient days, outpatient, accident and ER attendance, GP, practice nurses, community psychologists, psychiatric nurses and occupational therapist contacts	Interviews (Client Service Receipt Inventory)	Medical records	Retrospective - baseline, 6 and 12-months after trial entry
Goossens (2000) [14]	The Netherlands	RCT	Patients with chronic low back pain (N=40)	Physiotherapy contacts and specialist contacts	Cost diary	Health insurance company data	Prospective—each diary covered a period up to 4 weeks
Guerriere (2006) [15]	Canada	Sample of enrollees from a cystic fibrosis clinic	Patients with cystic fibrosis (N=110)	Home-based healthcare professional appointments, healthcare professional appointments outside home, laboratory and diagnostic tests, prescription medications	Questionnaire (The Ambulatory and Home Care Record)	Ontario Health Insurance Plan (OHIP) database, Ontario Home Care Administrative System (OHCAS) database, hospital pharmacy database	Daily basis completion over a 4-week data collection period
Hoogendoorn (2009) [16]	The Netherlands	RCT	Patients with chronic obstructive pulmonary disease (COPD) (N=175)	Daycare treatment, hospital admissions, hospital days, medical specialist visits, physiotherapist visits, respiratory nurse visits, dietician visits	Cost booklets	Care provider registrations	Prospective—each booklet covered a period of 4 weeks collected every 2 months

Reference & year	Country	Study design	Population (sample size)	Healthcare resource use	Method of self-report	Administrative record	Data collection period
Janson (2005) [17]	Sweden	RCT	Patients with potentially curable colon cancer (N= 40)	Outpatient consultations, doctor consultations, GP and district nurse, surgical ward readmissions	Telephone surveys	Clinical Case Record Forms (CRFs)	Retrospective—every 2 weeks with a 12-week follow-up period
Kennedy (2002) [18]	United Kingdom	RCT	Patients with a musculoskeletal condition (N=52)	Physiotherapy	Questionnaires	Outpatient department records (ODR), hospital physiotherapy department's records	Retrospective—3 and 12 months after initial outpatient appointment
Longobardi (2011) [19]	Canada	Population-based cohort	Adults with inflammatory bowel disease (IBD) (N=352)	Hospital and physician visits (IBD and non-IBD related)	Interview (standardized)	Health insurance provider data-bases	Retrospective—12 months
Marks (2003) [20]	USA	Sample of enrollees in a diabetes management program	Diabetics (N=1,230)	Hospitalizations, emergency room visits	Telephone assessment	Health insurance claims	Retrospective—6 months
Mirandola (1999) [21]	Italy	Sample of enrollees from the South-Verona Outcome Project	Patients in contact with the South-Verona Community Psychiatric Service (N=339)	Contacts with psychiatrists, psychologists, community and outpatient nurses, and social workers; admissions in sheltered accommodations; rehabilitation group contacts; day hospital contacts	Interview (Client Service Receipt Interview)	Psychiatric Case Register (PCR)	Retrospective—twice yearly

Reference & year	Country	Study design	Population (sample size)	Healthcare resource use	Method of self-report	Administrative record	Data collection period
Mistry (2005) [22]	United Kingdom	RCT	Patients consulting their GP regarding symptoms of depression (N=85)	Contacts with GPs and other practice staff, contacts with community mental health teams, psychiatric and general hospital contacts	Questionnaires	GP records (community-based or manual records)	Retrospective—3-month intervals over the course of 1 year
Patel (2005) [23]	England	Random sample of consecutive primary care attenders	Primary care attenders (N=229)	GP visits	Questionnaires (variant of the Client Service Receipt Inventory)	GP case records	Retrospective—6 months
Pinto (2011) [24]	New Zealand	RCT	Patients with hip or knee osteoarthritis (N= 50)	GP contacts, Total number of medical consultations	Questionnaire (Osteoarthritis Cost and Consequences Questionnaire)	Medical records	Retrospective—3 months
Pollicino (2002) [25]	Australia	RCT	Patients with apparently resectable non-small-cell lung cancer (N= 24)	Broad type of services (unreferred attendances, specialist attendances, diagnostic imaging, pathology tests)	Diaries	Health Insurance Commission (HIC)	Prospective—fortnightly the first year and monthly the year after
Richards (2003) [26]	United Kingdom	RCT	Acute ward patients aged 60 years or older (N=185)	Hospital admissions, visits by community healthcare staff	Interviewer-administered questionnaire	Healthcare provider reports (Patient Administration System, Integrated Community System, GP questionnaire)	Retrospective—12 weeks

Reference & year	Country	Study design	Population (sample size)	Healthcare resource use	Method of self-report	Administrative record	Data collection period
Van den Brink (2004) [27]	The Netherlands	Multi-center RCT	Patients with rectal cancer (N=147)	Medications, care products	Diary, questionnaire	Medical records (hospital admission system and pharmacy records)	Prospective (diary)—weekly from discharge to 3 months and monthly from 3 to 6 months / Retrospective (questionnaire) —3 and 6 months
Zentner (2012) [28]	Germany	Cluster RCT	People with severe mental illness (N=82)	Hospital admissions, ambulant psychiatric contacts, medication	Questionnaire (Client Sociodemographic and Service Receipt Inventory)	Health insurance administrative records (AOK)	Retrospective—6 months hospital admissions / 3 months ambulant contact / 1 month medication

MeRQ scores

Table 3 summarizes the methodological reporting quality as assessed by the reviewers using the MeRQ tool. The reporting of the elements was scored 'good', 'fair', 'poor', 'not reported', 'not addressed', or 'not applicable'. In seven of the sixteen studies, more than 75% of the criteria (≥ 9) were considered 'good' (explicit identification) [17-19, 21, 24, 26, 28]. Overall, the majority of studies scored at least 'fair' on most of the MeRQ criteria.

One study [14] did not explicitly identify its aim as evaluating the validity of self-reported measurements compared with administrative measurements. It was only at the end of the methods section that the aim of comparing both methods was mentioned. The deducibility was low, resulting in a poor reporting quality of the first criteria. Another study [13] did not address any validation arguments or strategies for interpreting the validation evidence. The same study lacked a comprehensive description of the measurement procedures for self-reports and administrative data. The only information provided was regarding the healthcare services that were measured and the time intervals between both collection methods. This study scored 'poor' on two quality criteria. Seven out of 16 studies did not report the time intervals between the self-reported data collection and the administrative data collection [16-22]. Pollicino et al. [25] did not report statistical methods for the assessment of validation estimates; they instead reported the difference counts between agreement and disagreement. Furthermore, no methods for statistical uncertainty were reported [25]. Five other studies also failed to report methods for statistical uncertainty [14, 16, 22, 27, 28]. Van den Brink et al. [27] calculated concordance between types of medication and care products reported using both methods. Percentages were provided in the figures and text; however, it was impossible to deduce the statistical methods used and the validation estimates reported. Although Mirandola et al. [21] reported a bias correction factor for precision, methods for statistical uncertainty were lacking, and the study scored 'poor' on this reporting element.

Table 3. MeRQ scores

Study	Aim to compare validation evidence	Strategies for inter-pretation validation evidence	Instru-ment eligibility criteria	Rationale for relation	Methods or proce-dures for self-reports	Methods or proce-dures for adminis-trative data	Time interval	Popula-tion charac-teristics	Statisti-cal methods for valida-tion estimates	Valida-tion evidence	Valida-tion estimates	Statisti-cal un-certainty	Conclu-sion and inter-pretation of valida-tion evidence
Byford (2007)	±	±	±	+	-	-	+	±	+	Agree-ment	+	+	+
Goos-sens (2000)	-	±	+	+	+	±	+	+	+	Validity	+	?	+
Guerrere (2006)	±	-	±	±	-	+	+	+	+	Agree-ment	+	+	+
Hoogen-doorn (2009)	±	±	±	+	+	+	?	+	+	Agree-ment	+	?	+
Janson (2005)	±	+	+	±	+	+	?	+	+	Agree-ment	+	+	+
Kennedy (2002)	+	±	+	+	+	+	?	+	+	Agree-ment	+	+	+
Longo-bardi (2011)	+	+	+	+	+	+	?	+	+	Validity	+	+	+
Marks (2003)	±	±	±	+	+	+	+	+	+	Agree-ment	+	+	±
Miran-dola (1999)	+	+	+	+	+	+	?	+	+	Agree-ment	+	-	+
Mistry (2005)	+	+	±	±	±	±	+	±	+	Agree-ment	+	+	±
Patel (2005)	±	±	-	+	±	±	+	±	+	Agree-ment	+	+	+
Pinto (2011)	+	+	+	+	+	+	+	+	+	Validity	+	+	+

Study	Aim to compare validation evidence	Strategies for interpreting validation evidence	Instrument eligibility criteria	Rationale for relation	Methods or procedures for self-reports	Methods or procedures for administrative data	Time interval	Population characteristics	Statistical methods for validation estimates	Validation evidence	Validation estimates	Statistical uncertainty	Conclusion and interpretation of validation evidence
Pollicino (2002)	+	?	+	+	-	-	+	-	?	Agreement	?	?	-
Richards (2003)	±	+	+	+	+	+	+	±	+	Agreement	+	+	+
Van Den Brink (2004)	±	+	±	+	+	+	+	±	+	Validity	+	±	±
Zentner (2012)	+	+	+	+	+	+	+	+	+	Validity	+	-	+

+ = good (explicit reporting in the study); ± = fair (reporting in study makes it possible to deduce what is expected); - = poor (not explicitly reported in the study and difficult to deduce what is expected); ? (not reported)

Validation evidence

The main statistical analysis and statistical uncertainty analyses, as well as the validation estimates were also captured during basic data abstraction and presented in Table 4.

The statistical methods used to calculate the healthcare resource use validation estimates varied widely. Most studies examined the correlation between self-reported and administrative measurements of healthcare resource use, with calculations including the Spearman's (rank) correlation coefficient [16, 17, 28], Pearson's correlation coefficient [19, 27], the interclass correlation coefficient [18], and primarily Lin's concordance correlation coefficient [13, 19, 21, 23, 24]. The Lin coefficient combines measurements of precision and accuracy to determine whether the data significantly deviate from perfect concordance. In an attempt to measure agreement, another four studies calculated a (weighted) kappa [15, 20, 22, 26]. Wilcoxon signed rank tests were used in one study to test differences in estimations of non-normal data distributions[14]. Agreement and proportion-perfect agreement were considered in the analyses of five studies [15, 16, 25-27]. In 56% of the studies, statistical uncertainty was defined. Most studies presented the 95% confidence intervals around the validation estimates [13, 15, 17, 18, 20, 24, 26]. The 95% limits of agreement were calculated in three studies [13, 23, 24].

Table 4. Validation evidence

Reference & year	Statistical analysis	Validation estimate (evidence)	Statistical uncertainty
Byford (2007) [13]	Lin concordance correlation coefficient (ρ_c)	Inpatient days (p_c 0.658) Outpatient appointments (p_c 0.231) Accident and ER attendance (p_c 0.760) General practitioner contacts (p_c 0.631) Practice nurse contacts (p_c 0.056) Community psychologist contacts (p_c 0.063) Community psychiatric nurse contacts (p_c 0.349) Occupational therapist contacts (p_c 0.001)	95% confidence interval 95% limits of agreement
Goossens (2000) [14]	Nonparametric Wilcoxon Signed Rank Test	No significant difference between the two methods for contacts with specialists (p-value 0.93) Significant difference for physiotherapist visits (p-value 0.021)	N.A.
Guerriere (2006) [15]	Observed agreement (Po) Simple Kappa (κ) Quadratic weighted Kappa (κ_w)	Physiotherapy (Po1.00 κ 1.00) Clinic visits (Po0.98 κ 0.90) Hospitalization (Po0.99 κ 0.88) Diagnostic tests (Po0.94 κ 0.78) Home-based nursing or personal support (Po0.97 κ 0.75) General practitioner visits (Po0.96 κ 0.69) Laboratory tests (Po0.92 κ 0.51) Medical specialist visits (Po0.85 κ 0.41) Diabetes therapy (Po0.95 κ_w 0.5) Antibiotics/antifungals (Po0.81 κ_w 0.63) Gastrointestinal therapy (Po0.78 κ_w 0.55) Pulmonary therapy (Po0.78 κ_w 0.5)	95% confidence interval
Hoogendoorn (2009) [16]	Spearman's rank correlation coefficient (ρ) Proportion of perfect agreement (p.a. %)	Daycare treatment (ρ 0.55 p.a. 87%) Daycare treatment for COPD (ρ 0.49 p.a. 98%) Hospital admissions (ρ 0.93 p.a.88%) Hospital admissions for COPD (ρ 0.94 p.a.95%) Total hospital days (ρ 0.91 p.a.79%) Total hospital days for COPD (ρ 0.93 p.a.93%) Medical specialist visits (ρ 0.70 p.a.8%) Physiotherapy (ρ 0.86 p.a.7%) Respiratory nurse visits (ρ 0.65 p.a.11%) Dietician visits (ρ 0.64 p.a.29%)	95% confidence interval
Janson (2005) [17]	Spearman's correlation coefficient (SCC)	Outpatient consultations doctor (SCC 0.66) General practitioner visits (SCC 0.49) District nurse visits (SCC 0.39) Surgical ward visits (SCC 0.72)	95% confidence interval
Kennedy (2002) [18]	Intraclass correlation coefficient	Physiotherapy (ICC 0.54)	95% confidence interval
Longobardi (2011) [19]	Concordance correlation coefficient (CCC) Pearson correlation coefficient (PCC)	Physician visits (CCC 0.58 PCC 0.68) Hospital overnights, IBD related (CCC 0.80 PCC 0.88) Hospital overnights, non-IBD related (CCC 0.86 PCC 0.93)	Bias Correction Factor (Cb)
Marks (2003) [20]	Kappa (κ)	Hospitalizations (κ 0.6366) Emergency room visits (κ 0.5390)	95% confidence interval

Reference & year	Statistical analysis	Validation estimate (evidence)	Statistical uncertainty
Mirandola (1999) [21]	Lin concordance correlation coefficient (ρ_c)	Psychiatrists community contacts (ρ_c 0.521) Psychiatrists outpatient contacts (ρ_c 0.538) Psychologists community contacts (ρ_c 0.495) Psychologists outpatient contacts (ρ_c 0.539) Nurses community contacts (ρ_c 0.45) Nurses outpatient contacts (ρ_c 0.94) Social workers community contacts (ρ_c 0.183) Social workers outpatient contacts (ρ_c 0.186) Admissions in sheltered accommodations (ρ_c 0.999) Rehabilitation group contacts (ρ_c 0.165) Day hospital contacts (ρ_c 0.140) Admissions to acute psychiatric ward (ρ_c 0.953)	N.A.
Mistry (2005) [22]	Weighted Kappa (kw)	General practitioner surgery contacts (kw 0.370) General practitioner home visits (kw 0.485) Psychiatric hospital and community service (kw 0.278) Social services (kw 0.021) Other hospital and specialist services (kw 0.430)	N.A.
Patel (2005) [23]	Lin concordance correlation coefficient (ρ_c)	General practitioner visits (ρ_c 0.756)	95% limits of agreement
Pinto (2011) [24]	Lin concordance correlation coefficient (ρ_c)	General practitioner contacts (ρ_c 0.41) Total number of medications (ρ_c 0.63)	95% limits of agreement 95% confidence interval
Pollicino (2002) [25]	Extent of agreement (absolute numbers, agreed on number of services used)	Unreferred visits (extent of agreement 3/23) Specialist visits (extent of agreement 9/22) Diagnostic imaging (extent of agreement 5/20) Pathology tests (extent of agreement 1/14)	N.A.
Richards (2003) [26]	Percentage crude agreement (%) (quadratic) Weighted Kappa (kw)	Hospital readmissions (85.7% kw0.42) General practitioner surgery visits (83.0% kw0.61) Patient-requested general practitioner home visits (41.7% kw0.59) Routine general practitioner home visits (71.0% kw0.71)	95% confidence interval
Van den Brink (2004) [27]	Percentage concordance	Medications and care products (25%)	N.A.
Zentner (2012) [28]	Spearman's rank correlation coefficient	Hospital admissions 0.53 Ambulant psychiatric contacts 0.33 Medications 0.31	N.A.
N.A. = not addressed			

Best evidence

In the end, six studies [18, 19, 21, 24, 26, 28] scored 'good' on all four MeRQ 'minimal criteria for reporting'. These studies were assessed and compared in detail. Three studies assessed the validity of the methods [19, 24, 28], whereas the other three studies assessed the methods based on their levels of agreement [18, 21, 26].

Longobardi et al. [19] assessed the validity of self-reported data gathered via standardized interviews compared with healthcare utilization drawn from an administrative database over a 12-month period for IBD patients. The authors concluded that there was good recall of medical contacts. However, systematic errors occurred in the reported frequency of physician visits (under-reporting by 35–45%) and overnight stays in the hospital (over-reporting by 25–35%) [19]. There was a reasonably good correlation of self-reported data with adequate concordance (0.59–0.86) and Pearson (0.68–0.93) correlation coefficients for both IBD and non-IBD related overnight stays in the hospital. Discrepancies in hospitalizations were due to failures to estimate a stay within the recall period, omitting stays following an emergency room visit, mistakenly including day surgery as hospitalization, and misclassifying an IBD stay. The tendency to over-estimate the number of overnight stays was not statistically significant, though a difference of this magnitude would likely have been significant in a larger sample with a higher base of hospitalizations. The self-reported data also underestimated the number of physician visits relative to the administrative data, despite the significant correlation. Regardless of whether IBD was inactive, fluctuating or active across the full 12-months, underestimation of the number of physician visits occurred by one-third when comparing self-reports with administrative data [19].

Pinto et al. [24] compared the validity of questionnaires and medical records regarding healthcare resource use over a 3-month period. Lin concordance correlation coefficients were fair to good, ranging from 0.41 for GP contacts to 0.63 for the number of medications reported. Generally, patients overestimated medication usage and underestimated GP contacts; however, the mean differences between the estimation periods were small: -0.22 for medication use and 0.06 for GP contacts. Participants more accurately reported their non-use than use of healthcare. GP contacts and medication usage had higher levels of sensitivity (93–100%) than specificity (63–100%). The three-month measurement period might explain this, as the use of healthcare might be relatively low over this short period of time. Compared cost estimations for GP contacts were in agreement with database results (0.502 Lin's concordance correlation coefficient).

Zentner et al. [28] assessed the validity of self-reported data gathered using a questionnaire compared with data drawn from health insurance administrative records over a six-month period. Correlation coefficients were moderate to low (0.53–0.31) and reflected statistical differences when comparing one measurement method to the other. Hospital admissions measured over a six-month recall period, ambulant psychiatric contacts measured over a three-month recall period and med-

ication use measured over a 1-month recall period were all underreported in the study. The low correlations between healthcare resource use measured using questionnaires and insurance registrations were significant, indicating a difference between the measurement methods [28].

Kennedy et al. [18] compared healthcare use measured via questionnaires with use measured via hospital physiotherapy department records. Agreement was measured by an intraclass correlation coefficient. The ICC of 0.54 (95% CI 0.31–0.70) reasonably supports the exchangeability of both the questionnaire and department records. However, self-reported data overestimated the resource use compared with hospital records; a mean of 1.3 (95% CI 0.4–2.2) fewer visits were reported in the physiotherapy department notes [18].

Mirandola et al. [21] assessed the agreement of self-reported healthcare use measured via interviews with psychiatric case registers over a six-month period (biannually). Twelve healthcare categories were defined, and agreement between self-reports and administrative data varied greatly for each category. Overall, agreement was particularly weak for day-patient (hospital) contacts (0.14) and very strong for nurses' outpatient contacts, number of acute psychiatric ward admissions and sheltered accommodation (all $\rho > 0.9$). These results translate into overall agreement (0.93) on the total costs. The findings support that self-reported and administrative measurements of healthcare use are exchangeable when estimating the total cost of services. Nonetheless, as the interview tends to overestimate the costs of day and community care (ρ 0.14 and ρ 0.41, respectively) and to underestimate outpatient costs (ρ 0.42), one should acknowledge the importance of adequate procedures for analyzing the complexity of costs components [21].

Richards et al. [26] assessed the agreement of self-reported data via interviewer-administered questionnaires compared with healthcare providers' reports over a 12-week period. Recall periods varied for different resources (4–12 weeks). Agreement and systematic differences in the distribution of paired responses were examined to assess the comparability of the methods. Within a range of 42–93%, agreement was over 72%. Systematic differences were observed when comparing self-reports with administrative data for routine primary care home visits (over-reporting). Underreporting of resources use occurred for hospital readmissions, requested primary care home visits, surgery consultations, district nursing and physiotherapy when compared with provider records over short timeframes (4–12 weeks). The number of visits was only gathered for hospital readmissions, routine and requested primary home care and surgery visits. Kappa coefficients were moderate or good for the four comparisons (kappa 0.42–0.86). The 95% confidence intervals of the kappa coefficients were quite wide for all comparisons. Systematic bias was indicated for hospital readmissions and surgery visits ($P < 0.05$). Marginally significant differences ($P = 0.05$ – 0.10) were indicated for requested and routine primary home care visits. There were adequate levels of crude agreement between patient-reported use of healthcare services and usage data from healthcare providers. However, the analysis of bias between pairs revealed systematic differences

and more marginally significant differences in the response patterns. Except for primary care routine home visits, patients underreported healthcare use compared with healthcare provider records [26].

Conclusions

This study draws attention to both the methodological reporting quality and the validation evidence of studies comparing self-reported with administrative data collection methods used in measuring healthcare resource use for cost estimation. Based on the small number of six studies with adequate methodological reporting quality presented in this review, the exchangeability of self-reported and administrative healthcare resource use measurements can only be cautiously supported based on the presented validation evidence (e.g. under- and over-reporting occurred, correlation coefficients were heterogeneous—including very weak ones, low correlations were presented, etc.). No evidence can be provided to suggest that one measurement is better than the other.

Throughout these six studies, when measured, patients generally tended to underestimate physician and physiotherapy contacts. The most commonly reported source of validation evidence was the calculation of correlation coefficients indicating the strength and direction of association. This approach is insufficient to indicate agreement or absolute measurement errors. Further attention should be paid to methods that enhance accuracy and to adequate methodological reporting quality. Notwithstanding, self-reported data appears to be able to provide adequate general estimates of healthcare use when administrative data are not available.

Limitations

A limitation of the study is the limited generalizability of the findings due to great diversity. Firstly, most studies collected data on multiple types of resource use and with divergent populations. However, the validation evidence was not necessarily more trustworthy when the population being assessed (e.g. patients with mental health complaints) was directly linked to the type of healthcare resource use relevant to the illness (e.g. psychologists). Secondly, as most studies were conducted in a randomized controlled trial, the main focus of these studies was on assessing the effectiveness and/or cost effectiveness of an intervention or policy adjustment. Comparing the self-reported data with administrative data was usually not the primary aim. Thirdly, sample sizes were small. In most cases, existing populations gathered for a completely different research purpose, were assessed in a secondary analysis on agreement levels.

Discussion

The methodological reporting quality was assessed with a newly developed appraisal tool called MeRQ, which was derived by combining items of existing guidelines, i.e. GRRAS and STARD. The MeRQ tool was developed to ensure the reporting of key elements required to assess the validation evidence and the risk of bias. It was difficult to appraise the methodological rigor when key information was reported vaguely or not at all. Interpreting validation evidence requires knowledge of the methods and statistics but even then, there are no standard statistics, and different interpretations are possible. Although the use of the MeRQ tool was feasible in the current systematic review, further research is warranted to assess its validity.

Adequate methodological reporting quality was found in a minority of the studies. Six studies assessed in the best evidence section demonstrate the complexity between different data sources. Data sets contained both random and systematic errors. Systematic errors were generally detected by comparing two imperfect sources of information, and taking directional trends into account might explain the discrepancies. Possible reasons for random sources of error could include problems in recall (for self-reports) or data entry errors (for administrative data). In both, interpretation or coding differences might explain random errors. Also, differential bias may occur in trials examining innovative services, particularly where the boundaries between specific service providers are blurred from the patient's perspective. These discrepancies in healthcare resource use that relate to cost elements might be of concern, especially when they have a large impact on the overall costs of care (e.g. hospital admissions, complex care packages). While the difference between two methods of data collection might be statistically significant, they might not be economically relevant, or vice versa. For example, if the healthcare resource use being researched (e.g. physiotherapy contacts) was one of the dominant costs in the evaluation (i.e. due to the population-specific health complaints), then the addition or subtraction of one extra visit per patient may not be statistically significant but would be economically relevant. Therefore, one needs to properly address a priori usage and identification of items for making cost estimations.

Implications for research and practice

The assumptions made in some studies that administrative databases should be the reference standard is questionable, because they seem to have errors. When using the data for cost estimations, statistical corrections might be relevant. However, corrections for cost data from administrative databases when computing incremental cost effectiveness ratios are so far unavailable [29]. Previous reviews identified limitations of these databases [3, 5, 6]. The health economic community does not yet have a culture of validating resource use measurement, and there is a lack of

CHAPTER 3

evidence regarding the collection of resource use data [30]. The MeRQ tool is a suitable instrument for further assessing the methodological reporting quality in order to deduce the validation evidence. The validation evidence of a (small) number of studies with adequate methodological reporting quality cautiously supports the exchangeability of both the self-reported and administrative resource use measurements. Coherent methodological quality which supports the interpretations of estimates and rigorous, well-reported research providing strategically collected validation evidence is needed.

References

1. Drummond M, McGuire A. *Economic Evaluation in Health Care: Merging theory with practice* New York: Oxford University Press; 2001.
2. Hakkaart L, Tan SS, Bouwmans, CAM. *Manual for cost research. Methods and standards costs for economic evaluations in healthcare* [in Dutch: Handleiding voor kostenonderzoek. Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg]. Rotterdam, the Netherlands: Instituut voor Medical Technology Assessment, Erasmus Universiteit Rotterdam, 2010.
3. Bhandari A, Wagner T. Self-reported utilization of health care services: improving measurement and accuracy. *Medical care research and review* : MCRR 2006;63:217-35.
4. Ridyard CH, Hughes DA. Development of a database of instruments for resource-use measurement: purpose, feasibility, and design. *Value in health* 2012;15:650-5.
5. Evans C, Crawford B. Patient self-reports in pharmaco-economic studies. Their use and impact on study validity. *Pharmacoeconomics* 1999;15:241-56.
6. Harlow SD, Linet MS. Agreement between questionnaire data and medical records. The evidence for accuracy of recall. *Am J Epidemiol* 1989;129:233-48.
7. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009;62:1006-12.
8. Sassi F, Archard L, McDaid D. Searching literature databases for health care economic evaluations: how systematic can we afford to be? *Medical care* 2002;40:387-94.
9. Alton V, Eckerlund I, Norlund A. Health economic evaluations: how to find them. *International journal of technology assessment in health care* 2006;22:512-7.
10. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, Roberts C, Shoukri M, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;64:96-106.
11. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC, Standards for Reporting of Diagnostic A. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003;326:41-4.
12. Streiner DL, Kottner J. Recommendations for reporting the results of studies of instrument and scale development and testing. *J Adv Nurs* 2014;
13. Byford S, Leese M, Knapp M, Seivewright H, Cameron S, Jones V, Davidson K, Tyrer P. Comparison of alternative methods of collection of service use data for the economic evaluation of health care interventions. *Health economics* 2007;16:531-6.
14. Goossens ME, Rutten-van Molken MP, Vlaeyen JW, van der Linden SM. The cost diary: a method to measure direct and indirect costs in cost effectiveness research. *J Clin Epidemiol* 2000;53:688-95.
15. Guerriere DN, Ungar WJ, Corey M, Croxford R, Tranmer JE, Tullis E, Coyte PC. Evaluation of the ambulatory and home care record: Agreement between self-reports and administrative data. *Int J Technol Assess Health Care* 2006;22:203-10.
16. Hoogendoorn M, van Wetering CR, Schols AM, Rutten-van Molken MP. Self-report versus care provider registration of healthcare utilization: impact on cost and cost utility. *International journal of technology assessment in health care* 2009;25:588-95.
17. Janson M, Carlsson P, Haglund E, Anderberg B. Data validation in an economic evaluation of surgery for colon cancer. *Int J Technol Assess Health Care* 2005;21:246-52.
18. Kennedy AD, Leigh-Brown AP, Torgerson DJ, Campbell J, Grant A. Resource use data by patient report or hospital records: do they agree? *BMC health services research* 2002;2:2.
19. Longobardi T, Walker JR, Graff LA, Bernstein CN. Health service utilization in IBD: comparison of self-report and administrative data. *BMC health services research* 2011;11:137.
20. Marks AS, Lee DW, Slezak J, Berger J, Patel H, Johnson KE. Agreement between insurance claim and self-reported hospital and emergency room utilization data among persons with diabetes. *Disease management* : DM 2003;6:199-205.
21. Mirandola M, Bisoffi G, Bonizzato P, Amaddeo F. Collecting psychiatric resources utilisation data to calculate costs of care: a comparison between a service receipt interview and a case register. *Soc Psychiatry Psychiatr Epidemiol* 1999;34:541-7.

CHAPTER 3

22. Mistry H, Buxton M, Longworth L, Chatwin J, Peveler R. Comparison of general practitioner records and patient self-report questionnaires for estimation of costs. *The European journal of health economics : HEPAC : health economics in prevention and care* 2005;6:261-6.
23. Patel A, Rendu A, Moran P, Leese M, Mann A, Knapp M. A comparison of two methods of collecting economic data in primary care. *Family practice* 2005;22:323-7.
24. Pinto D, Robertson MC, Hansen P, Abbott JH. Good agreement between questionnaire and administrative databases for health care use and costs in patients with osteoarthritis. *BMC Med Res Methodol* 2011;11:45.
25. Pollicino C, Viney R, Haas M. Measuring health system resource use for economic evaluation: a comparison of data sources. *Aust Health Rev* 2002;25:171-8.
26. Richards SH, Coast J, Peters TJ. Patient-reported use of health service resources compared with information from health providers. *Health & social care in the community* 2003;11:510-8.
27. van den Brink M, van den Hout WB, Stiggelbout AM, van de Velde CJ, Kievit J. Cost measurement in economic evaluations of health care: whom to ask? *Med Care* 2004;42:740-6.
28. Zentner N, Baumgartner I, Becker T, Puschner B. Health service costs in people with severe mental illness: patient report vs. administrative records. *Psychiatr Prax* 2012;39:122-8.
29. Johnston K, Buxton MJ, Jones DR, Fitzpatrick R. Assessing the costs of healthcare technologies in clinical trials. *Health Technol Assess* 1999;3:1-76.
30. Thorn JC, Coast J, Cohen D, Hollingworth W, Knapp M, Noble SM, Ridyard C, Wordsworth S, Hughes D. Resource-use measurement based on patient recall: issues and challenges for economic evaluation. *Applied health economics and health policy* 2013;11:155-61.

Appendix 1. MeRQ Tool: Criteria of Methodological Reporting Quality as Adapted from the Standards for Reporting Diagnostic Accuracy (STARD) and the Guidelines for Reporting Reliability and Agreement Studies (GRRAS)

Study and reviewer identification		
Reviewer:	First author:	Year of publication:
Title:		
Language:		Country:
Methodological Reporting Quality (MeRQ)		
Study criterion and reporting element (operational definition)	In this study, reporting of the criterion is:	
Explicit identification of the study as an evaluation of the reliability, agreement or validity (accuracy) of a self-reported healthcare resource use measurement instrument compared with healthcare resource use measured using administrative data	Good Fair Poor	Not reported Not addressed Not applicable
Proposed validation arguments and strategies for interpreting validation evidence	Good Fair Poor	Not reported Not addressed Not applicable
The measurement methods (eligibility criteria) with a critical review of evidence relevant for the study	Good Fair Poor	Not reported Not addressed Not applicable
Rationale for relationship between the self-report and the administrative data	Good Fair Poor	Not reported Not addressed Not applicable
Methods or measurement procedures for the self-reports	Good Fair Poor	Not reported Not addressed Not applicable

Methods or measurement procedures for the administrative data	Good Fair Poor	Not reported Not addressed Not applicable
Time interval between self-reports and administrative data collection	Good Fair Poor	Not reported Not addressed Not applicable
Population characteristics (including sample size, sampling methods, recruitment)	Good Fair Poor	Not reported Not addressed Not applicable

Please choose 1 of the following boxes to fill in according to the information provided in the article: Does this study examine: validity? reliability? agreement? (fill in according to your assessment)

Validity

Statistical methods for validation estimates defined	Good Fair Poor	Not reported Not addressed Not applicable
Validation estimates defined	Good Fair Poor	Not reported Not addressed Not applicable
Methods for statistical uncertainty defined (Variability - Confidence Intervals)	Good Fair Poor	Not reported Not addressed Not applicable

Reliability

Statistical methods for validation estimates defined	Good Fair Poor	Not reported Not addressed Not applicable
--	----------------------	---

Validation estimates defined	Good Fair Poor	Not reported Not addressed Not applicable
Methods for statistical uncertainty defined (Variability - Confidence Intervals)	Good Fair Poor	Not reported Not addressed Not applicable
Agreement		
Statistical methods for validation estimates defined	Good Fair Poor	Not reported Not addressed Not applicable
Validation estimates defined	Good Fair Poor	Not reported Not addressed Not applicable
Methods for statistical uncertainty defined (Variability - Confidence Intervals)	Good Fair Poor	Not reported Not addressed Not applicable
Concluding validation argument / Interpretation of validation evidence	Good Fair Poor	Not reported Not addressed Not applicable
Additional information		

CHAPTER 3

Appendix 2. Search strategy. Database: Medline (PubMed)

(measure*OR correlat* OR evaluat* OR accuracy OR accurate OR precision OR agreement OR valid*[Title/Abstract]) **AND** (self report [MeSH Terms] OR self-report [Text Word] OR self-assessment[Text Word] OR patient self report [Text Word] OR patient-reported [Text Word] OR questionnaires [Mesh Term] OR recall [Text Word] OR interviews as topic [Mesh Term] OR health care surveys [Mesh Terms]) **AND** (Medical Records [MeSH Terms] OR Records as Topic [Mesh Term] OR medical records systems, computerized [Mesh Term], OR medical record administrators [Mesh Term]) **AND** (health resources [MeSH Terms] OR delivery of health care [MeSH Terms] OR utilization [Text Word] OR resource use [Text Word]) **AND** data collection [Mesh Term]) **AND** cost analysis

CHAPTER

4

Design of a trial-based economic evaluation on the cost effectiveness of employability interventions among work disabled employees or employees at risk of work disability: The CASE-study

Cindy Noben
Frans Nijhuis
Angelique de Rijk
Silvia Evers

Abstract

Background: In the Netherlands, absenteeism and reduced productivity due to work disability lead to high yearly costs reaching almost 5% of the gross national product. To reduce the economic burden of sick leave and reduced productivity, different employability interventions for work-disabled employees or employees at risk of work disability have been developed. Within this study, called 'CASE-study' (Cost effectiveness Analysis of Sustainable Employability), five different employability interventions directed at work disabled employees with divergent health complaints will be analyzed on their effectiveness and cost effectiveness. This paper describes a consistent and transparent methodological design to do so.

Methods/Design: Per employability intervention 138 participants are needed whereof approximately 69 participants receiving the intervention will be compared with 69 participants receiving usual care. Based on the intervention-specific characteristics, a randomized control trial or a quasi-experiment with match-criteria will be conducted. Notwithstanding the study design, eligible participants will be employees aged 18 to 63, working at least 12 hours per week, and at risk of work disability, or already work-disabled due to medical restrictions. The primary outcome will be the duration of sick leave. Secondary outcomes are health status and quality of life. Outcomes will be assessed at baseline and then six, 12 and 18 months later. Economic costs will consist of healthcare costs and cost of lost production due to work disability, and will be evaluated from a societal perspective.

Discussion: The CASE-study is the first to conduct economic evaluations of multiple different employability interventions based on a similar methodological framework. The cost effectiveness results for every employability intervention will be published in 2014, but the methods, strengths and weaknesses of the study protocol are discussed in this paper. To contribute to treatment options in occupational health practice and enable the development of guidelines on how to conduct economic evaluation better suited to this field; this paper provides an important first step.

Trial Registration: Four trials involved in the CASE-study are registered with the Netherlands Trial Registry: Care for Work (NTR2886), Health and Motion (NTR3111), Guidance to Excel in Return to Work (NTR3151), Care for Companies/Second Care (NTR3136).

Keywords: economic evaluation, cost effectiveness, employability, return-to-work, work disability.

Published as: C. Y. Noben, F. J. Nijhuis, A. E. de Rijk, S. M. Evers. Design of a trial-based economic evaluation on the cost effectiveness of employability interventions among work disabled employees or employees at risk of work disability: The CASE-study. *BMC Public Health* 2012;12:43.

Background

Work disability

Many recent studies have focused on reduced productivity and work absence due to work disability in Western countries [1-5]. Work disability prevents or detracts the employee's productivity and has many economic and public health consequences, such as problems for employers, employees, their families, and society [6-8]. In Western countries its main consequences are financial, due to reduced productivity, increased job turnover and the cost of additional rehabilitation programs [9-11]. In order to estimate the costs of work disability, staff absence and presence needs to be recorded. Absenteeism refers to the total days lost from work. Presenteeism refers to attending work whilst still sick or disabled, causing reduced productivity while at work [12]. Absenteeism and presenteeism can be used as integrated measures of physical, psychological and social functioning in studies of working populations [12, 13]. In the Netherlands, the Sickness Absence Reduction Act was introduced in 1994 in an attempt to reduce absenteeism. Although there has been a major reduction in absenteeism, nearly three million Dutch workers still suffer from one or more chronic conditions which constrains their work performance [14]. Sickness absence remains a large problem in terms of costs, labor participation and social consequences [11, 15, 16]. The most common causes of sickness absence are: musculoskeletal complaints, mental health problems, and cardiovascular diseases [17, 18]. Chronic diseases are more common among workers of 45 years and older, and it is expected that this age group will greatly increase, leading to a growing number of occupational health problems [17, 18]. In the Netherlands, the total annual costs of absenteeism arising from health complaints and work disability amounted to 26 billion Euros, which accounts for almost 5% of the gross national product [14]. Additionally, Kremer and Steenbeek (2010) have computed that approximately 21.5% of absenteeism from work could be avoided if health services and employer and employee opportunities improved. To do this, shorter referral times, swifter medical care, improved collaboration between health care providers, reduced workload and an improved work-life balance are important factors [17, 18].

Return-to-Work

Different return-to-work (RTW) interventions have been developed and evaluated to reduce sickness absence due to work disability. The effect, and occasionally the cost effectiveness, of RTW interventions have been demonstrated in several observational studies [19-24], randomized controlled trials [25-30] and reviews [31, 32]. For example, Franche et al. [31] concluded that work-based RTW interventions among employees with work-disability caused by musculoskeletal or other pain-

related conditions can reduce work disability duration and associated costs. Lambeek et al. [33] showed that an integrated care program for work-disabled participants with chronic low back pain has the potential to significantly reduce societal costs and improve care effectiveness, quality of life and function on a broad scale [33]. Other studies demonstrate that implementing preventive interventions before sickness absence occurs can be successful in preventing long-term sickness absence [19, 29, 34, 35]. Whilst these results are very promising for society, participants, and employers, more research is necessary to confirm the long-term effects and cost effectiveness of such interventions. However, methodological limitations in the current literature make it difficult to assess the feasibility of the different interventions proposed. These limitations include restriction of the population of interest due to the strict inclusion criteria, restrictions in the nature of the intervention evaluated, restrictions in study-design and the measurement of outcomes. There is moreover, a lack on transparency and consensus about how to conduct trial-based economic evaluations within the general field of employability.

Objectives

This paper is the first of its kind to describe a transparent, uniform methodological design for the economic evaluation of several different trial-based interventions in the field of employability, disability and return-to-work. Known as the ‘CASE-study’ (Cost effectiveness Analysis of Sustainable Employability), the project’s aim is to evaluate at least five different employability interventions. All of the interventions share the common aim of reducing absenteeism and optimizing the employability of employees with a known disability or those at risk of disability due to musculoskeletal complaints, mental health problems or a combination of both musculoskeletal and mental health complaints. In contrast with previous studies, the project design has no population restrictions and allows for the economic evaluation of any employability intervention regardless of the population they were developed for. Furthermore, the project also includes interventions provided in the primary and specialist health sectors as well as those delivered within the workplace. It also includes interventions aimed at reducing health care costs, wage replacement costs and other intervention costs in addition to improvements in return to disability duration and work return rates as measured by self-reported return to work dates, total duration of absence and subsequent recurrences, etc. Other outcomes include quality of life, quality of work life and general health status, etc. In addition, we cast a wide net to include a range of study designs since RCTs are not always possible in the economic evaluation of employability interventions. This paper describes the patient inflow in the event of an RCT or a quasi-experiment with a matched controlled design being conducted. For every employability intervention, an effect study will be conducted. In addition, from a societal perspective, a trial-based economic evaluation for every employability intervention will be carried out in a consistent manner. The results of these analyses will be discussed elsewhere in the

future, but we hypothesize that the employability intervention is preferable in terms of costs, effects and utility compared to usual care. This paper will therefore demonstrate the strength of the CASE-study project in prescribing a methodological design which allows for the assessment of several different types of employability interventions within varying populations.

Methodology

Methods/design

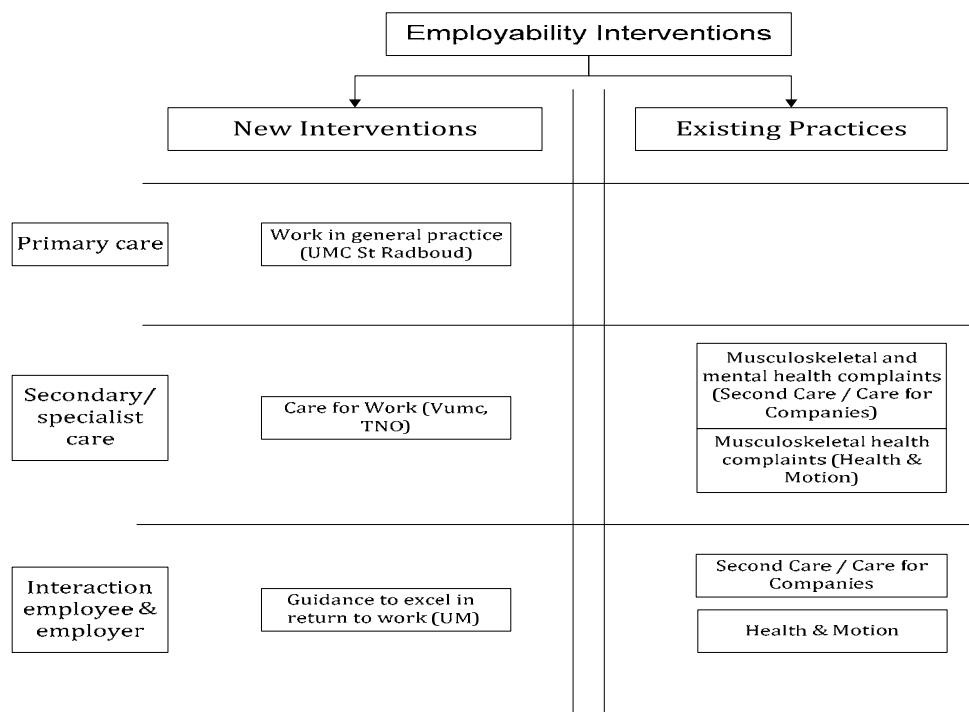
Organization of the CASE-study

We will carry out an economic evaluation designed for at least five Dutch employability interventions. Those undertaking the evaluation are not connected with the organizations or individuals who have developed, or are in the process of developing the employability interventions so no competing interests arise. Within the CASE-study each existing or innovative intervention will be compared with usual care for work-disabled employees or those at risk of work disability due to medical restrictions. At the start of the CASE-study, nine programs were interested in participating in the project. Unfortunately four programs could not collaborate because they were unable to provide the minimum required number of participants eligible for the study, or could not provide a comparable control group. The remaining five employability interventions have the ability to comply to the minimal requirements and focus on primary care, secondary or specialist care, and/or on the interaction between the employer and employee. Two existing employability interventions (called 'Best Practices') and three newly developed employability interventions are included in the study. The employability intervention focusing on primary care aims to acknowledge the importance of employability in general practice. This intervention is under development at the University of Nijmegen (UMC St Radboud). The employability intervention in specialist care is being developed by the VU University Medical Centre Amsterdam (VUmc) and focuses on employability among patients who suffer from rheumatoid arthritis. The third newly developed intervention aims at improving the interaction between the employee and his/her employer by appointing an independent mediator. The intervention is being developed by a separate team at Maastricht University. One of the 'Best Practice' interventions, developed by Health & Motion Nederland, offers patients with physical health complaints the ability to implement a range of exercises learned at the physiotherapeutic setting in their working environment by means of a workplace-intervention. The fifth and final intervention is developed by Second Care/Care for Companies (Best Practice) and aims at achieving sustainable employability among patients with physical and/or psychological health complaints using a multidisciplinary approach. An overview of these interventions is provided in Figure 1.

It is beyond the scope of this design article to describe the precise content of each employability intervention in detail. This will be done when the results of the economic evaluation per employability intervention are published. What is of importance in this paper is that the employability interventions use an innovative, systematic approach in order to support work participation in employees with health complaints. These are innovative because both the healthcare system and the private/work life of the employees are considered in the intervention. The features of the interventions differ according to the patient population, but the aim remains the same: to reduce sickness absence and increase the potential of achieving sustainable employability.

The protocol for this study has been approved by the Ethics Committee of Psychology (ECP107) at Maastricht University, the Netherlands.

Figure 1. Employability interventions included in the CASE-study



Usual work resumption care in the Netherlands

The Gatekeeper Improvement Act stipulates that the responsibility for work resumption lies with the employer and the employee. The employer is obliged to start an employee’s work resumption program as quickly as possible, so the employee can resume his/her own work or other adequate work. In return, the employee should accept the work activities the employer provides. Furthermore, and unique in the world, the Dutch reintegration process provides the employer with salary and mandatory reintegration obligations for a period of two years [36]. Participants included in the care as usual group will continue to receive care for their health complaints, but attention to work-related factors influencing these will only occur minimally. Nevertheless, since attention to labor and health-related characteristics is obligatory under the Gatekeeper Improvement Act [36], participants in the care as usual group who are work-disabled will also receive adequate care (care as usual). Participants included in the control group will be treated and supervised according to good clinical practice based on the evidence-based guidelines of the Dutch association of occupational physicians [37, 38].

Participants

Employees eligible for participation in the study will be work-disabled or at risk of work disability due to medical restrictions. In order to identify these employees, we will use employer registration forms, registrations with health care organizations and insurance companies, and registration forms from those providing the intervention. Employees eligible for inclusion will be 18 to 63 years of age, perform paid labor for at least 12 hours per week, and are able to communicate in Dutch. Reasons for exclusion are absenteeism for 1.5 years or longer, fulltime students with a student job, informal caregivers and/or volunteers.

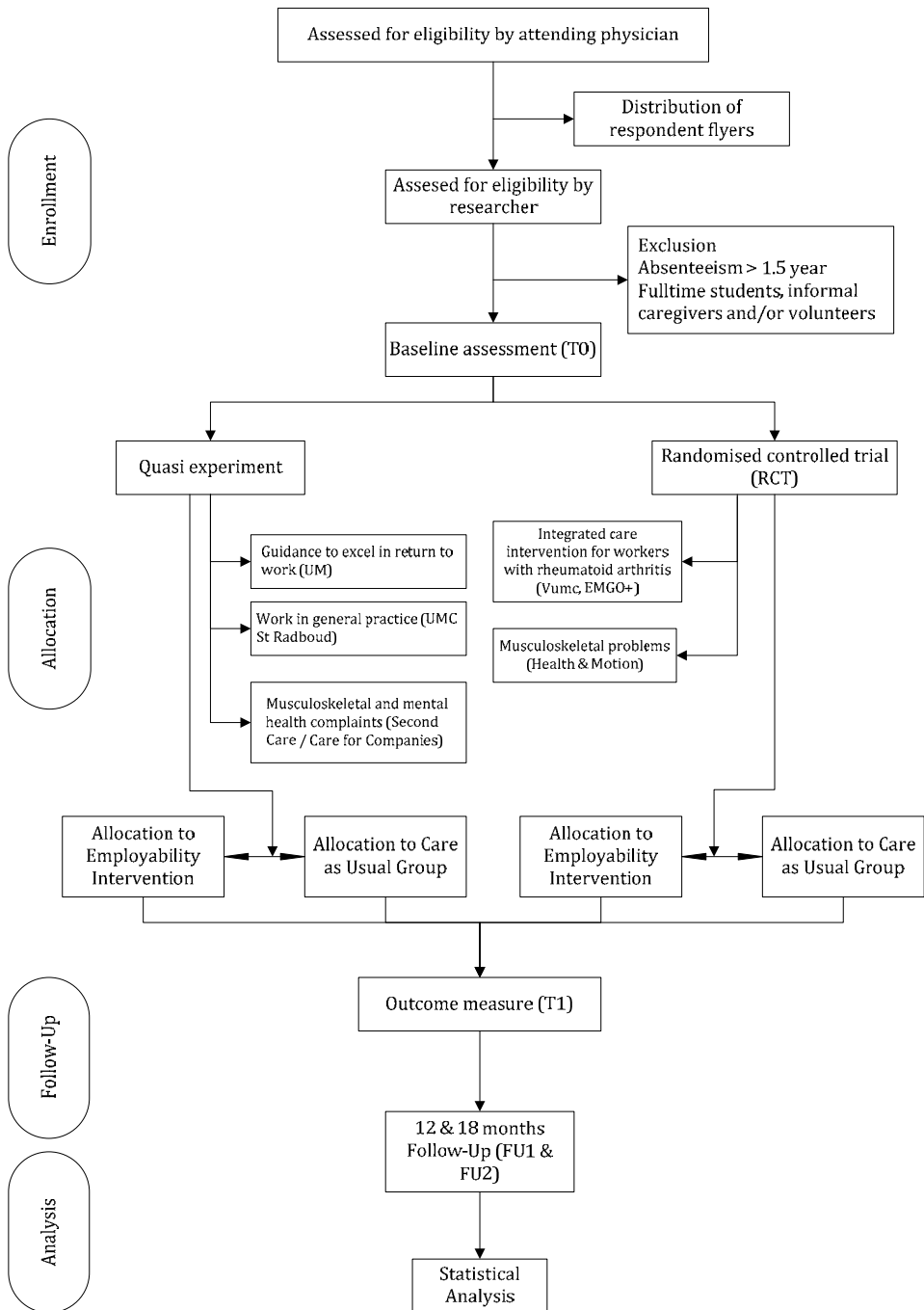
In order to recruit a sufficient number of eligible employees, one or more health care providers providing an employability intervention will screen service users for inclusion in the study. When appropriate, information on the research will be provided to the patient. This will be done orally by the health care provider, possibly supported by a handout containing contact information for the researchers, what is expected of participants, privacy concerns, and so on. A self-completion card will be attached to each flyer (handout) consisting of a short checklist for inclusion and informed consent form for signing. The employee is requested to fill in the short checklist, sign the informed consent form, and send these back to the researcher. Participants will have at least one week to consider their participation in the study before signing the informed consent form. Participants will be screened for exclusion by the researcher on the basis of answers provided on the checklist. If an employee meets all the selection criteria and is willing to sign the consent form, randomization or matching will then place. Before the intervention starts, the first (baseline) measurement occurs. An email with a login code is sent to all participants, and they will be asked to fill in the baseline questionnaire. The questionnaire for the intervention group is identical to the one for the control group.

Study design

As suggested by Drummond et al. (2001), a randomized controlled trial (RCT) remains the standard design for evaluating the relative effect of candidate interventions [39]. However, due to practical reasons, randomly assessing participants to the employability intervention or care as usual is not always possible in the CASE-study. Employers and occupational physicians are legally obliged to offer sickness absence guidance, and to do so, they refer employees to one or more employability interventions. Thus, a quasi-experiment, comparing employees who are allocated to an experimental intervention in a non-random manner with those non-randomly assigned to a conventional intervention is regarded as the next best study design. Unlike an RCT, a quasi-experiment consists of a control group derived from a large non-equivalent population. The variables are not manipulated, and participants are selected on the basis of matched criteria with those in the non-intervention group. Randomization is not used therefore, and participants/control members are selected on the basis of their having an equal chance of demonstrating a causal relationship between the intervention and its outcomes for the control or experimental

intervention. Both quasi-experiments and RCTs are acceptable for the economic evaluation of employability interventions. The anticipated flow of subject enrolment in a quasi-experiment or RCT is shown graphically in Figure 2.

Figure 2. Participants flow



Randomization / matching

In the case of randomization, the minimization method will be applied. Minimization aims to ensure an excellent balance for several prognostic factors in small groups [40]. Employees consenting to participate will be pre-stratified according to age (18-32 years, 33-48 years, 49-64 years), gender, and working hours (12-24 hours/week, >24 hours/week). Since health care providers will treat participants in both the intervention and in the control groups, the study cannot be blinded. Matching i.e. formulating experimental and control groups on the basis of pairing participants on the basis of predetermined matching criteria will be used. These matched-criteria are age, working hours and health complaints. Thus, an employee in the intervention group will be matched to someone meeting the same criteria who will then act as a member of the control group.

Timeline of the study

During recruitment, employees will be asked whether they want to participate in the research by one or more health care providers operating in the employability intervention. Participation requires employees to fill in a self-reported online questionnaire at time points T0 (baseline), T1 (six months after admission), FU1 (12 month follow-up) and FU2 (18 month follow-up). After eighteen months the study results for sustainable employability and quality of life are expected to extend into subsequent years. The timeline of the study can also be found in Figure 2.

Outcome assessment and data collection

The outcome assessment and data collection is identical for every employability intervention. Questions derived from different questionnaires, such as the 'National Working Conditions Survey' (NEA) [18], 'EuroQol 5 Dimensions 5 Levels' (EQ-5D-5L) [41], 'Short Form Health Survey' (SF-36) [42], 'Trimbos/iMTA questionnaire for costs associated with psychiatric illness' (TIC-P) [43], and 'the productivity and disease questionnaire' (PRODISQ) [44], are combined into a retrospective self-reported questionnaire. The economic evaluation will be analyzed from a societal perspective. This means that the most appropriate set of costs captured from the data (e.g. health care costs, productivity costs etc.), regardless of where the costs or benefits occur, will be applied [36].

Effect evaluation*Primary outcome measure*

The primary outcome measure in the CASE-study is reduced sickness absence duration in calendar days. Reduced sickness absence will be measured subjectively via patient self-report records. The duration of sick leave will be calculated as the time in calendar days since reduced productivity occurred (baseline). This will then be

compared with the productivity time of a person at six, 12, and 18 months. The success of an intervention will be defined by comparing the average reduction in sickness absence between the intervention group and the control group. The employability intervention is effective when the employees in the intervention group return to work twice as quickly as employees in the control group. Data on sickness absence duration will be gathered by using the productivity and disease questionnaire (PRODISQ) [46].

Secondary outcome measures

Data on secondary outcome measures will be gathered by means of self-administered questionnaires such as, NEA [18], EQ-5D-5L [41], SF-36[42], TIC-P [43] and PRODISQ[44].

Sustainable employability

After commencing a treatment intervention, a follow-up measurement will take place at 12 and 18 months to analyze the sustainability of employability. An intervention will be deemed to be unsuccessful when absenteeism of more than fourteen consecutive days after receiving the productivity intervention occurs. On the basis of previous research, it will be assumed that the long-term effect of the employability intervention stabilizes after 18 months [24].

Self-rated health

Self-related health will be assessed using one item from the Short Form Health Survey (SF-36). This widely used measurement tool describes how a respondent values and assess his/her general health [45]. Data gathered via questions derived from the NEA [18] will be used to detect health complaints underlying work disability and co-morbidity.

Quality of Life

An explorative analysis of productivity costs and quality of life showed a clear relationship between both variables [45]. Therefore, quality of life will be measured in work-disabled employees by using the EuroQoL five dimensions, five levels (EQ-5D-5L) [41]. The EQ-5D-5L comprises 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. There are five possible answers for each dimension, defining (5⁵) 3,125 possible health states. Furthermore, the EQ-5D-5L includes a visual analogue scale (VAS) rating from zero (worst imaginable health state) to 100 (best imaginable health state) [41].

Economic evaluation

Self-reported questionnaires will capture the costs and benefits from a societal perspective, thus independently of those who bear the costs, and those who receive the benefits. Self-reported questionnaires appear to be a reliable alternative for cost-diaries, at least for a recall period of six months, since participants will still remember significant life events [46-49]. Costs that arise in different years will be discounted at 4% [50]. Costs directly related to the provision of the employability

intervention will be objectively registered via the providers of the employability intervention. Direct healthcare costs and direct non-healthcare costs comprise respectively: costs of visits for primary and secondary care, home care, drugs, and for visits to an alternative or informal health therapist. Direct costs will be measured by an adapted version of the Trimbos/iMTA questionnaire for costs associated with psychiatric illness [43] and will be estimated on the basis of prices suggested in the cost calculation guidelines for healthcare in the Netherlands [50]. The prices of prescribed drugs will be based on Daily Defined Dosage (DDD) taken from the Royal Dutch Society for Pharmacy [50]. Indirect costs will be determined by measuring data on the costs of lost productivity due to work disability and are derived through the PRODISQ [45]. The indirect costs of production loss due to sick leave will be calculated by means of the friction cost method whereby the number of days of sickness absence and lost earnings are used to calculate the costs of production losses. Lost earnings will be determined on the basis of the differences between income at baseline and income at follow-up. Based on a mean added value of the Dutch working population, production losses are confined to the period needed to replace a sick employee [44]. The indirect costs of productivity losses without sick leave will be based on the number of days in which productivity was hindered due to health complaints and an estimation of the efficiency in these days [51]. Furthermore, compensation will be taken into account when determining production loss. Information about the number of hours to catch up with the work will be obtained as well as the way in which this compensation was achieved. If these data are considered invalid, for example because the income data of too many employees are unknown, we will use national data of average production value per worker, related to the education level and labor sector.

Sample size

For the economic evaluation of the three newly developed employability interventions, the study sample size was pre-specified by the researchers and the developers of the particular interventions. For the sample size calculation of the existing 'Best Practices', we expect that a Hazard Ratio (HR) of 2.0 indicates a relevant difference in the average number of hours/days of lost work due to work disability between employees in the intervention group and employees in the control group after 18 months. This HR is based on findings of recent studies in occupational health care [25, 52-55]. With a power of 80% at a 0.05 significance level, the sample size was calculated to be 120 employees [56]. Taking into account a dropout rate of 15%, 5% per follow-up measurement time, a total of 138 participants are needed. A dropout is defined as an employee who has not filled in the questionnaire at one or more measurement time-points.

Data analyses

Effectiveness

All statistical analyses, performed to distinguish differences between the control group and the intervention group, will be performed according to the intention-to-treat principle. This means that the participants will remain in the group they were allocated to at baseline. All data regarding primary and secondary outcomes will be analyzed using descriptive statistics. To control for protocol deviations likely to cause bias, data will be analyzed using per-protocol analyses whereby outcomes will be compared between those participants who received at least one or a partial intervention session and the control group. With the aim of correcting for possible regression to the mean, the baseline value of the particular outcome variable will be added to the model. Days of sick leave in the year before the inclusion, measured at baseline, will be added as a potential effect-modifier.

Diagnosis and prognostic parameters can be confounders, thus the Cox regression analysis for confounders will be used to adjust for differences between control and intervention groups. If necessary, differences in baseline will be adjusted for prognostic dissimilarities. Missing data will be handled using SPSS missing value analysis on item level. Completely missing measurements will be handled using multiple imputation.

The primary dependent variable is days of absence from work. In order to estimate the number of absence days, survival analyses will be used. There are two reasons why we opted to use survival statistics. Firstly, the time to sustainable employability is unlikely to be normally distributed which will make parametric statistical analysis difficult or inappropriate. Increasing productivity can occur quite early, possibly within months, but can also, as expected in the care as usual situation, may not be prolonged or may even decrease again. Secondly, survival analysis accounts for dropout by the technique of censoring. Participants who drop out will have the same hazard of sustainable employment than those that remained in the study. The Cox Proportional hazard model will therefore be used to analyze the HR of the days of sick leave duration. Reduced sickness absence duration in calendar days will be analyzed with the Kaplan Meier survival method, and the differences in absence duration between both groups will be tested with a log rank test.

To analyze the secondary outcomes, longitudinal multilevel analyses will be used to assess the differences in health and quality of life between the control group and the intervention group. Data will be analyzed in SPSS 17.0.

Economic evaluation

The economic evaluation will be carried out according to the intention-to-treat principle. Non-parametric bootstrapping (1,000 replications) will be used to obtain 95% confidence intervals to estimate the uncertainty surrounding the cost differences, because cost-data are usually skewed to the right [57].

For the cost effectiveness analysis, the incremental cost effectiveness ratio (ICER) will be calculated by assessing the ratio of the differences in costs between the groups to the differences in effects between the groups. The ICER indicates whether the additional investments needed for the interventions gain at least one extra unit of effect compared with care as usual [57]. The cost utility ratio will be calculated by the incremental costs divided by the difference in QALYs [39, 58].

Bootstrapping with 5,000 replications will be used to estimate the sample-uncertainty around the point estimates (ICER and ICUR). The bootstrapped ICERs and ICURs will be graphically presented on a cost effectiveness plane. Furthermore, acceptability curves will show the probability that the employability intervention is cost effective at a specific ceiling ratio, which is the maximum amount of money society is willing to pay to gain one extra unit of effect or a gain in quality of life [59]. Finally, the robustness of the different parameters on the cost effectiveness calculation will be assessed by sensitivity analyses. Data processing will be done in SPSS 17.0 and bootstrapping will be carried out in Excel.

Discussion

This study protocol provides detailed information on how to assess the effectiveness and cost effectiveness of employability interventions among employees with work disabilities or employees at risk of becoming work-disabled due to health complaints. The results derived from these economic evaluations can benefit the individual employee, the employer and the broader occupational health practice sector because illness duration, care consumption and costs of lost productivity are expected to decrease if the employability intervention is cost effective. For society as a whole, cost effective employability interventions are attractive since they are expected to increase labor participation and public health. Furthermore, we feel that productivity is the key to higher economic growth as we are facing an ageing population.

A limitation of this study-design is the generalizability to other countries. The employability interventions are specifically tailored for the Dutch context in which they will be implemented. When applying these interventions in other countries, specific characteristics of the population in relation to the societal, political and cultural context in which the interventions are to be used should be taken into account. Although international generalizability might be restricted and require adaptation to the intervention, lessons can be learned and implementation advice given to those engaging in future work.

Selection bias cannot be ruled out in this design because we ask the healthcare provider to select employees to participate in our study. As a consequence, healthcare providers cannot be blinded. Furthermore, the consultation with the health care provider might change the participants' point of view on work disabil-

ity. This ‘Hawthorne effect’ will be born in mind and might overestimate or underestimate the effect of the intervention. However, we feel that minor interactions such as the provision of information about the study will not be an important source of bias.

The strengths of this pragmatic design looking at the cost effectiveness of employability interventions versus usual care, is the comparison under ‘real life’ circumstances of the intervention and control variables. Resource use is self-reported by participants and will be collected separately from the unit costs or prices of those resources. We will keep in mind the possibility that health care and intervention providers in different settings may wish to apply their own prices to the units of resource use. Furthermore, due to the standard recall period of six months, an adequate estimation of resource use is linear over time. This linear time trend restricts overestimation and underestimation of resource use between two consecutive measurement moments. An additional strength of this design is the inclusion of losses due to sick leave and the costs of using health care facilities together with productivity losses during working periods whilst unwell (presenteeism). In developing the methodology, we kept in mind the possibility that the employability and usual care interventions may have different effects on the component of reduced productivity while at work and therefore capture these. Besides internal validity, external validity of the results of the economic evaluation is reached by analyzing the data from a societal perspective. We can split the calculated cost effectiveness from the societal perspective to other perspectives, e.g. employee perspective, organizational perspective, etc.

To conclude, the research design provided in this article will be applied to each employability intervention included in the CASE-study. The CASE-study is the first study assessing the cost effectiveness of different employability interventions. In addition, this design might contribute to new and better guidelines in the conduction of future economic evaluations in the occupational health sector. The results of the CASE-study will become available between 2012 and 2014.

List of abbreviations

CASE-study	Cost effectiveness Analysis Sustainable Employability
CEA	Cost effectiveness Analysis
CUA	Cost Utility Analysis
DDD	Daily Defined Doses
ECP	Ethics Committee of Psychology
EQ-5D-5L	EuroQol 5Dimensions 5Levels
HR	Hazard Ratio
ICER	Incremental Cost Effectiveness Ratio
ICUR	Incremental Cost Utility Ratio
NEA	National working conditions survey
PRODISQ	PROductivity and DISease Questionnaire
QALY	Quality of Life
RCT	Randomized Control Trial
RTW	Return to Work
SF-36	Short form (36) health survey
TIC-P	Trimbos/iMTa questionnaire for Cost associated with Psychiatric Illness
VAS	Visual Analogue Scale

Acknowledgements

To create national support for the study, collaboration is established between the different providers of the employability interventions, Maastricht University and the School for Public Health and Primary Care (CAPHRI). The providers of the employability interventions agreed with the conditions as included in the collaboration contract. Within this study there is also collaboration between Maastricht University, KPMG Plexus (office for consultancy and research management) and TNO (office for research and innovation). The CASE-study is part of a national project called 'Work and Care' (http://www.instituutgak.nl/research_programs/arbeid-en-gezondheidszorg/) and is funded by the Foundation Institute Gak (SIG). We would also like to thank Dr. Stephen J. O'Connor for proof reading and making final amendments to the completed manuscript in respect of its readability.

References

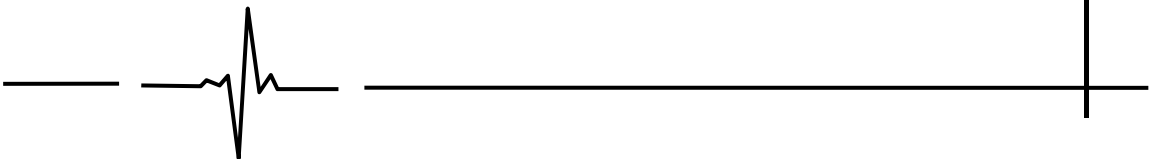
1. Maki K, Vahtera J, Virtanen M, Elovainio M, Pentti J, Keltikangas-Jarvinen L, et al. Sickness absence among female employees with migraine and co-existing conditions. *Cephalalgia*. 2008 Nov;28(11):1136-44.
2. Ervasti J, Kivimaki M, Pentti J, Suominen S, Vahtera J, Virtanen M. Sickness absence among Finnish special and general education teachers. *Occup Med*. 2011 Aug 23.
3. Melchior M, Ferrie JE, Alexanderson K, Goldberg M, Kivimaki M, Singh-Manoux A, et al. Does sickness absence due to psychiatric disorder predict cause-specific mortality? A 16-year follow-up of the GAZEL occupational cohort study. *Am J Epidemiol*. 2010 Sep 15;172(6):700-7.
4. Melchior M, Ferrie JE, Alexanderson K, Goldberg M, Kivimaki M, Singh-Manoux A, et al. Using sickness absence records to predict future depression in a working population: prospective findings from the GAZEL cohort. *Am J Public Health*. 2009 Aug;99(8):1417-22.
5. Kivimaki M, Head J, Ferrie JE, Singh-Manoux A, Westerlund H, Vahtera J, et al. Sickness absence as a prognostic marker for common chronic conditions: analysis of mortality in the GAZEL study. *Occup Environ Med*. 2008 Dec;65(12):820-6.
6. Ferrie JE, Vahtera J, Kivimaki M, Westerlund H, Melchior M, Alexanderson K, et al. Diagnosis-specific sickness absence and all-cause mortality in the GAZEL study. *J Epidemiol Community Health*. 2009 Jan;63(1):50-5.
7. Lund T, Kivimaki M, Christensen KB, Labriola M. Socio-economic differences in the association between sickness absence and mortality: the prospective DREAM study of Danish private sector employees. *Occup Environ Med*. 2009 Mar;66(3):150-3.
8. Birnbaum HG, Kessler RC, Kelley D, Ben-Hamadi R, Joish VN, Greenberg PE. Employer burden of mild, moderate, and severe major depressive disorder: mental health services utilization and costs, and work performance. *Depress Anxiety*. 2010;27(1):78-89.
9. Virtanen P, Siukola A, Luukkaala T, Savinainen M, Arola H, Nygard CH, et al. Sick leaves in four factories--do characteristics of employees and work conditions explain differences in sickness absence between workplaces? *Scand J Work Environ Health*. 2008 Aug;34(4):260-6.
10. Severens JL, Mulder J, Laheij RJ, Verbeek AL. Precision and accuracy in measuring absence from work as a basis for calculating productivity costs in The Netherlands. *Soc Sci Med*. 2000 Jul;51(2):243-9.
11. Henderson M, Glozier N, Holland Elliott K. Long term sickness absence. *BMJ*. 2005 Apr 9;330(7495):802-3.
12. Lofland JH, Pizzi L, Frick KD. A review of health-related workplace productivity loss instruments. *Pharmacoeconomics*. 2004;22(3):165-84.
13. Marmot M, Feeney A, Shipley M, North F, Syme SL. Sickness absence as a measure of health status and functioning: from the UK Whitehall II study. *J Epidemiol Community Health*. 1995 Apr;49(2):124-30.
14. Statistics Netherlands. National statistics on sick leave, frequency, period of absence 2010 [13 September 2011]; Available from: <http://www.statline.cbs.nl>.
15. Watson PJ, Main CJ, Waddell G, Gales TF, Purcell-Jones G. Medically certified work loss, recurrence and costs of wage compensation for back pain: a follow-up study of the working population of Jersey. *Br J Rheumatol*. 1998 Jan;37(1):82-6.
16. Williams DA, Feuerstein M, Durbin D, Pezzullo J. Health care and indemnity costs across the natural history of disability in occupational low back pain. *Spine (Phila Pa 1976)*. 1998 Nov 1;23(21):2329-36.
17. Kremer AM, Steenbeek R. Avoidable sickness absence in a Dutch working population. *J Occup Rehabil*. 2010 Mar;20(1):81-9.
18. Jehoel-Gijsbers G. Beter aan het werk. Trendrapportage ziekteverzuim, arbeidsongeschiktheid en werkhervatting. Den Haag: Sociaal en Cultureel Planbureau 2010.
19. Post M, Krol B, Groothoff JW. Work-related determinants of return to work of employees on long-term sickness absence. *Disabil Rehabil*. 2005 May 6;27(9):481-8.

20. Anderzen I, Arnetz BB. The impact of a prospective survey-based workplace intervention program on employee health, biologic stress markers, and organizational productivity. *J Occup Environ Med.* 2005 Jul;47(7):671-82.
21. Franche RL, Severin CN, Hogg-Johnson S, Cote P, Vidmar M, Lee H. The impact of early workplace-based return-to-work strategies on work absence duration: a 6-month longitudinal study following an occupational musculoskeletal injury. *J Occup Environ Med.* 2007 Sep;49(9):960-74.
22. Mushet GR, Miller D, Clements B, Pait G, Gutterman DL. Impact of sumatriptan on workplace productivity, nonwork activities, and health-related quality of life among hospital employees with migraine. *Headache.* 1996 Mar;36(3):137-43.
23. Meerding WJ, Looman CW, Essink-Bot ML, Toet H, Mulder S, van Beeck EF. Distribution and determinants of health and work status in a comprehensive population of injury patients. *J Trauma.* 2004 Jan;56(1):150-61.
24. Young AE, Roessler RT, Wasiaik R, McPherson KM, van Poppel MN, Anema JR. A developmental conceptualization of return to work. *J Occup Rehabil.* 2005 Dec;15(4):557-68.
25. van Oostrom SH, Anema JR, Terluin B, de Vet HC, 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.
26. Uegaki K, Bakker I, de Bruijne M, van der Beek A, Terluin B, van Marwijk H, et al. Cost effectiveness of a minimal intervention for stress-related sick leave in general practice: results of an economic evaluation alongside a pragmatic randomised control trial. *J Affect Disord.* 2010 Jan;120(1-3):177-87.
27. van Oostrom SH, Heymans MW, de Vet HC, van Tulder MW, van Mechelen W, Anema JR. Economic evaluation of a workplace intervention for sick-listed employees with distress. *Occup Environ Med.* 2010 Sep;67(9):603-10.
28. van Oostrom SH, van Mechelen W, Terluin B, de Vet HC, Knol DL, Anema JR. A workplace intervention for sick-listed employees with distress: results of a randomised controlled trial. *Occup Environ Med.* 2010 Sep;67(9):596-602.
29. Brouwers EP, Tiemens BG, Terluin B, Verhaak PF. Effectiveness of an intervention to reduce sickness absence in patients with emotional distress or minor mental disorders: a randomized controlled effectiveness trial. *Gen Hosp Psychiatry.* 2006 May-Jun;28(3):223-9.
30. Lambeek LC, van Mechelen W, Knol DL, Loisel P, Anema JR. Randomised controlled trial of integrated care to reduce disability from chronic low back pain in working and private life. *BMJ.* 2010;340:c1035.
31. Franche RL, Cullen K, Clarke J, Irvin E, Sinclair S, Frank. Workplace-based return-to-work interventions: a systematic review of the quantitative literature. *J Occup Rehabil.* 2005 Dec;15(4):607-31.
32. Tompa E, Dolinschi R, de Oliveira C. Practice and potential of economic evaluation of workplace-based interventions for occupational health and safety. *J Occup Rehabil.* 2006 Sep;16(3):375-400.
33. Lambeek LC, Bosmans JE, Van Royen BJ, Van Tulder MW, Van Mechelen W, Anema JR. Effect of integrated care for sick listed patients with chronic low back pain: economic evaluation alongside a randomised controlled trial. *BMJ.* 2010;341:c6414.
34. Kant IJ, Jansen NW, van Amelsvoort LG, Swaen GM, van Leusden R, Berkouwer A. Screening questionnaire Balansmeter proved successful in predicting future long-term sickness absence in office workers. *J Clin Epidemiol.* 2009 Apr;62(4):408-14 e2.
35. Kant I, Jansen NW, van Amelsvoort LG, van Leusden R, Berkouwer A. Structured early consultation with the occupational physician reduces sickness absence among office workers at high risk for long-term sickness absence: a randomized controlled trial. *J Occup Rehabil.* 2008 Mar;18(1):79-86.
36. Hop AD, Pijper de B, Sens IM, Wildenburg A, Baars MJH, Dekker EM, et al. De kleine gids voor de Nederlandse sociale zekerheid 2010. twee 81 ed. Hop AD, editor. Hoofddorp: Kluwer; 2010.
37. Rebergen D, Hoenen J, Heinemans A, Bruinvels D, Bakker A, van Mechelen W. Adherence to mental health guidelines by Dutch occupational physicians. *Occup Med (Lond).* 2006 Oct;56(7):461-8.
38. van der Klink JJ, van Dijk FJ. Dutch practice guidelines for managing adjustment disorders in occupational and primary health care. *Scand J Work Environ Health.* 2003 Dec;29(6):478-87.

39. Drummond M, McGuire A. *Economic Evaluation in Health Care: Merging theory with practice*. New York: Oxford University Press; 2001.
40. Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials. a review. *Control Clin Trials*. 2002 Dec;23(6):662-74.
41. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011 April.
42. Jenkinson C, Coulter A, Wright L. Short form 36 (SF36) health survey questionnaire: normative data for adults of working age. *BMJ*. 1993 May 29;306(6890):1437-40.
43. Hakkaart-van Roijen L, van Straten A, Donker M, Tiemens B. Manual Trimbos/iMTA questionnaire for Costs associated with Psychiatric illness (TiC-P) [In Dutch: Handleiding Trimbos/iMTA questionnaire for Costs associated with Psychiatric illness (TiC-P)]: Rotterdam: Instituut voor Medische Technology Assessment. 2002.
44. Koopmanschap MA. PRODISQ: a modular questionnaire on productivity and disease for economic evaluation studies. *Expert Rev Pharmacoecon Outcomes Res*. 2005 Feb;5(1):23-8.
45. Brouwer WB, Meering WJ, Lamers LM, Severens JL. The relationship between productivity and health-related QOL: an exploration. *Pharmacoeconomics*. 2005;23(3):209-
46. van den Brink M, van den Hout WB, Stiggelbout AM, Putter H, van de Velde CJ, Kievit J. Self-reports of health-care utilization: diary or questionnaire? *Int J Technol Assess Health Care*. 2005;21(3):298-304.
47. van den Brink M, van den Hout WB, Stiggelbout AM, van de Velde CJ, Kievit J. Cost measurement in economic evaluations of health care: whom to ask? *Med Care*. 2004 Aug;42(8):740-6.
48. Pole JD, Franche RL, Hogg-Johnson S, Vidmar M, Krause N. Duration of work disability: a comparison of self-report and administrative data. *Am J Ind Med*. 2006 May;49(5):394-401.
49. Laestadius JG, Ye J, Dimberg L. Can we trust the answers? Reliability and validity of self-reported sick leave due to musculoskeletal symptoms. *J Occup Environ Med*. 2008 Jun;50(6):611-3.
50. Hakkaart L, Tan SS, Bouwmans, CAM. Manual for cost research. Methods and standards costs for economic evaluations in healthcare [in Dutch: Handleiding voor kostenonderzoek. Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg]. Rotterdam, the Netherlands: Instituut voor Medical Technology Assessment, Erasmus Universiteit Rotterdam, 2010.
51. Brouwer WB, Koopmanschap MA, Rutten FF. Productivity losses without absence: measurement validation and empirical evidence. *Health Policy*. 1999 Jul;48(1):13-27.
52. Lambeek LC, Anema JR, van Royen BJ, Buijs PC, Wuisman PI, van Tulder MW, et al. Multidisciplinary outpatient care program for patients with chronic low back pain: design of a randomized controlled trial and cost effectiveness study. *BMC Public Health*. 2007;7:254.
53. Anema JR, Steenstra IA, Bongers PM, de Vet HC, Knol DL, Loisel P, et al. Multidisciplinary rehabilitation for subacute low back pain: graded activity or workplace intervention or both? A randomized controlled trial. *Spine (Phila Pa 1976)*. 2007 Feb 1;32(3):291-8; discussion 9-300.
54. Steenstra IA, Anema JR, Bongers PM, de Vet HC, van Mechelen W. Cost effectiveness of a multi-stage return to work program for workers on sick leave due to low back pain, design of a population based controlled trial. *BMC Musculoskelet Disord*. 2003 Nov 21;4:26.
55. Rebergen DS, Bruinvels DJ, van der Beek AJ, van Mechelen W: Design of a randomized controlled trial on the effects of counseling of mental health problems by occupational physicians on return to work: the CO-OP-study. *BMC Public Health*. 2007;7:183.
56. Power and sample size program Available from: http://linus.nci.nih.gov/brb/sample-size/sdpap_survival.html.
57. Barber JA, Thompson SG. Analysis of cost data in randomized trials: an application of the non-parametric bootstrap. *Stat Med*. 2000 Dec 15;19(23):3219-36.
58. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost effectiveness analysis: the importance of controlling for baseline utility. *Health Econ*. 2005 May;14(5):487-96.
59. Drummond M, McGuire A. *Economic Evaluation in Health Care: Merging theory with practice* New York: Oxford University Press; 2001.

PART II

EXISTING PRACTICES



CHAPTER

5

Comparative cost effectiveness of two interventions to promote work functioning by targeting mental health complaints among nurses: Pragmatic cluster randomized trial

Cindy Noben
Filip Smit
Karen Nieuwenhuijsen
Sarah Ketelaar
Fania Gärtner
Brigitte Boon
Judith Sluiter
Silvia Evers

Abstract

Background: The specific job demands of working in a hospital may place nurses at elevated risk for developing distress, anxiety and depression. Screening followed by referral to early interventions may reduce the incidence of these health problems and promote work functioning.

Objective: To evaluate the comparative cost effectiveness of two strategies to promote work functioning among nurses by reducing symptoms of mental health complaints. Three conditions were compared: the control condition consisted of online screening for mental health problems without feedback about the screening results. The occupational physician condition consisted of screening, feedback and referral to the occupational physician for screen-positive nurses. The third condition included screening, feedback, and referral to e-mental health.

Design: The study was designed as an economic evaluation alongside a pragmatic cluster randomised controlled trial with randomisation at hospital-ward level.

Setting and participants: The study included 617 nurses in one academic medical centre in the Netherlands.

Methods: Treatment response was defined as an improvement on the Nurses Work Functioning Questionnaire of at least 40% between baseline and follow-up. Total per-participant costs encompassed intervention costs, direct medical and non-medical costs, and indirect costs stemming from lost productivity due to absenteeism and presenteeism. All costs were indexed for the year 2011.

Results: At six months follow-up, significant improvement in work functioning occurred in 20%, 24% and 16% of the participating nurses in the control condition, the occupational physician condition and the e-mental health condition, respectively. In these conditions the total average annualised costs were €1,752, €1,266 and €1,375 per nurse. The median incremental cost effectiveness ratio for the occupational physician condition versus the control condition was dominant, suggesting cost savings of €5,049 per treatment responder. The incremental cost effectiveness ratio for the e-mental health condition versus the control condition was estimated at €4,054 (added costs) per treatment responder. Sensitivity analyses attested to the robustness of these findings.

Conclusions: The occupational physician condition resulted in greater treatment responses for less costs relative to the control condition and can therefore be recommended. The e-mental health condition produced less treatment response than the control condition and cannot be recommended as an intervention to improve work functioning among nurses.

Keywords: cost effectiveness, mental disorders, nurses, occupational health, prevention & control, randomised controlled trial, stress, work functioning.

Published as: Noben C, Smit F, Nieuwenhuijsen K, et al. Comparative cost effectiveness of two interventions to promote work functioning by targeting mental health complaints among nurses: pragmatic cluster randomised trial. *Int J Nurs Stud.* 2014 Oct;51(10):1321-1331.

Introduction

Nurses are at elevated risk for mental distress, anxiety and depression [1-4]. Possible explanations for this increased risk are found in work characteristics such as high job demands and a lack of autonomy [2,5] Poor mental health is not only undesirable in its own right, but will likely also have an adverse impact on the nurses' job functioning and may thus jeopardise the health and safety of the patients in their care. After all, nurses with poor mental health experience significantly more medical errors [2,6,7]. For these reasons it is imperative to protect and promote mental health in nurses, and to monitor and safeguard the quality of their functioning at work [2].

Mental disorders carry substantial disease and economic burdens. Preventive interventions for mental disorders exist; however, what interventions should be financed and implemented is an issue that needs to be addressed by decision makers. Moreover, the number of health-economic evaluations that were conducted in the work setting is very limited. Likewise, information to aid in the transferability of available results to different contexts and settings is limited [8]. Economic evaluations can provide answers, select interventions that are cost effective and avoid wasting limited resources. An approach to priority setting is largely based on economic techniques to assess the cost effectiveness to answer questions regarding the economic value for money of competing interventions [9,10].

Periodic screening might be useful to identify nurses with signs of mental health problems and encourage help-seeking behaviour. To that end a Worker Health Surveillance was developed. The Worker Health Surveillance is a preventive strategy that aims at the early detection of negative health effects at work [2,11,12]. A Worker Health Surveillance with personalised feedback and referral to dedicated early interventions for screen positives might be a successful strategy to prevent the onset and further deterioration of mental health problems and to reduce impairments in work functioning [2,13]. In this study we compare a control condition consisting of screening without feedback versus Worker Health Surveillance screening with feedback plus referral for a consultation with an occupational physician or referral to preventive e-mental health interventions. These approaches have not been evaluated from a health-economic perspective.

Therefore, the aim of this study is to assess the comparative cost effectiveness of the occupational physician condition and the e-mental health condition versus the control condition, with a view to protecting mental health and improving and sustaining work functioning in nurses.

Methods

Design

The *Mental Vitality @ Work* study [14] was designed as a pragmatic cluster randomised controlled trial, with randomisation at the level of hospital wards to three conditions:

- 1) Screening and feedback followed by referral to the occupational physician for screen-positives (the occupational physician condition).
- 2) Screening and feedback followed by referral and access to preventive e-mental health interventions (the e-mental health condition).
- 3) Screening without feedback and without referral to either the occupational physician or the e-mental health interventions (the control condition).

Data were recorded at baseline and after three and six months. In the economic evaluation, we assessed the comparative cost effectiveness in two contrasting scenarios: (1) the occupational physician condition versus the control condition, and (2) the e-mental health condition versus the control condition. A medical ethics committee approved the study.

Randomization

Cluster randomisation was performed at the ward level to prevent contamination between participants working in the same ward. A pre-randomisation procedure with incomplete-double-consent design was applied [15,16] meaning that individuals were only informed about their own group. This further minimised the possibility of contamination.

Randomisation was conducted using the computer software programme Nquery Advisor in blocks of three wards. After randomisation, 28 wards with 591 employees were assigned to the occupational physician condition, 29 wards with 579 employees to the e-mental health condition, and 29 wards with 561 employees to the control-condition.

Sample

The study population of the complete trial included all nurses with similar work demands and work conditions; including surgical nurses, anaesthetic nurses, and allied health professionals (henceforth: *nurses*) working in one Dutch academic medical centre. Nurses who were sick-listed at the start of the study and expected to be on sick leave for more than two weeks were excluded from the study. All eligible employees were invited to take part in the study, which for the complete trial added up to 1,731 employees working in 86 wards. Cost effectiveness analyses were performed according to the intention-to-treat principle for the whole sample.

However, participation rates of the nurses in the interventions at baseline were 34% in the control condition; 32% in the occupational physician condition; and 31% in the e-mental health condition.

After randomisation, 212 nurses were assigned to the e-mental health condition, 210 to the occupational physician condition and 211 to the control condition. Sixteen nurses (three in the occupational physician condition, eight in the e-mental health condition and five in the control condition) were sick-listed for more than two weeks at the start of the trial, did not contribute to the data and were excluded from the analysis. Thus, the study population comprised a total of 617 nurses: 207 in the occupational physician condition, 204 in the e-mental health condition and 206 in the control condition.

Interventions

All participants were screened for work functioning impairments and six types of mental health complaints: distress, work-related fatigue, risky drinking, depression, anxiety, and post-traumatic stress disorder. Nurses in the control condition filled out the questionnaires and no further steps were taken. After completing the screening, the occupational physician condition and the e-mental health condition immediately received personalised feedback about their screening results.

In the occupational physician condition, screening and feedback were followed by an invitation for the screen-positives to attend the occupational physician. The nurses consulted the occupational physician or not at their own discretion. In order to structure the consultation of the occupational physician, a seven-step protocol was applied, with the focus on identifying impairments in work functioning and providing advice on how to improve wellbeing and work functioning.

In the e-mental health condition, screening and feedback were followed by referral to e-mental health interventions. The e-mental health interventions offered in the e-mental health condition were *Psyfit*, aimed at promoting mental fitness and wellbeing; *Strong at Work*, aimed at learning skills to cope better with work-related stress; *Colour your Life*, for coping with depressive symptoms; *Don't Panic Online*, to reduce symptoms of panic disorder; and *Drinking Less*, aimed at reducing risky alcohol consumption. Nurses who screened positive on one of these health problems were offered access to the corresponding e-mental health intervention. Nurses screening negative on mental health complaints, but positive on work functioning impairments were offered *Psyfit* and an onscreen psycho-educational leaflet about dealing with these impairments. Nurses screening negative on both mental health complaints and on work functioning impairments were only offered free access to *Psyfit*. The onscreen psycho-educational leaflet was also offered when nurses screened positive on mental health complaints and on work functioning impairments. In any case, making use of the e-mental health interventions was strictly voluntary and nurses were free to reject the offer of using the interventions [14].

Outcome measure

The primary outcome was work functioning as measured by the following subscales of the Nurses Work Functioning Questionnaire: cognitive aspects of task execution, causing incidents at work, avoidance behaviour, conflicts and irritations with colleagues, impaired contact with patients and their family, lack of energy and motivation. The Nurses Work Functioning Questionnaire is a 50-item self-report questionnaire with Likert-type response scales ranging from 0 (*totally disagree*) to 6 (*totally agree*); 0 (*disagree*) to 4 (*agree*); and 0 (*no difficulty*) to 6 (*great difficulty*) [11]. Internal reliability of the *Nurses Work Functioning Questionnaire* is high, with Cronbach's alphas varying between 0.70 and 0.94 [17].

The difference between the occupational physician condition and the e-mental health condition versus the control condition was examined as the percentage of individuals who improved at follow-up. The primary outcome of work functioning is operationalised as job-specific impairments in work functioning and were measured using a total score of the Nurses Work Functioning Questionnaire. The minimal important change value for improvement was based on the relative pre-post change scores, that is $(T_0 - T_1 / T_0) \times 100\%$, indicating the percentage of change on impaired work functioning in relation to the baseline score. Individuals with a relative improvement on their Nurses Work Functioning Questionnaire total score of 40% or more [18], which is the minimal important change (MIC) value, were defined as *relevantly improved*, and henceforth denoted as treatment responders.

Resource use and costing

Resource usage and costs entailed can be split into (a) intervention costs; (b) direct medical costs (due to health service uptake and pharmacy use); (c) direct non-medical costs (the nurses' out-of-pocket costs for travel and parking, incurred while making use of health services); and (d) indirect costs stemming from lost productivity in paid work due to absenteeism and presenteeism. All costs were in Euro indexed for the reference year 2011 based on the price indices from Statistics Netherlands [19].

Intervention costs

When calculating the intervention costs, a distinction was made between human and material resources. Material resources are divided into capital items that have a time span longer than one year, such as the equipment that was needed to build the screening module for the provision of the feedback and the interventions, and recurrent or revenue items that are consumed in less than one year, such as maintenance costs. For human resources, costs were based on the valuation of the personnel involved in the development and the application of the intervention. The corresponding calculations for the intervention costs can be obtained from the first author.

The per-participant costs for the online *Worker Health Surveillance* were estimated to be €3.80. The costs of the occupational physician intervention consisted of the per-participant screening costs of €3.80 plus the costs of the occupational physician at €73.11 per contact. For the e-mental health intervention, the per-participant costs were the screening costs of €3.80 plus the costs of the specific e-mental health interventions. The costs of the e-mental health interventions were only charged when a participant logged in to an e-mental health intervention and thus became a user of that intervention. The e-mental health interventions have fixed per-participant cost prices that are based on their market values, which are subject to change over time. In the reference year of 2011, the per-user costs were as follows: *Psyfit* €30, *Strong at Work* €175, *Colour your Life* €195, *Don't Panic Online* €225 and *Drinking Less* €45.

Direct medical costs

Health service costs were calculated by multiplying the health service units (contact, session, hour) with their standard full economic cost price. The standard costs were reported in the Dutch guideline for health economic evaluations [20] and indexed for the year 2011 using the consumer price index from Statistics Netherlands [19]. The costs of prescription drugs were calculated as the price per standard daily dose as reported in Dutch guidelines and multiplied by the number of days [21]. The pharmacist's dispensing costs of €5.99 and the general practitioners' prescription costs of €14 were added [20]. Over the counter drugs were based on their market prices.

Direct non-medical costs

The participants' travel and parking expenses incurred in receiving professional help were computed as the distance to a health service multiplied by the costs per kilometre (€0.21), with parking costs (€3 per hour) added [20].

Indirect non-medical costs

Finally, the costs stemming from production losses in paid work were calculated from the number of days absent from work (absenteeism) plus the number of workdays lost due to work cutback (presenteeism). Presenteeism was calculated by correcting for the degree of inefficiency, resulting in an inefficiency score used as point prevalence for the calculation of presenteeism costs. The inefficiency score for work quantity and work quality derived from items from the Productivity and Disease Questionnaire (ranging from 0 to 1, with 0 meaning not inefficient and 1 completely inefficient) was multiplied by the number of days at work while not feeling well in order to compute the costs of presenteeism [22].

The valuation method for productivity is rooted in the human capital theory whereby the production losses are assumed to equal the present value of all lost future earnings of the individual. That is, income (before tax) acts as a proxy for the production value of that individual and encompasses all productivity losses by this

person [23,24]. The costs of productivity losses were then assessed by multiplying the number of workdays lost by the gender and age-specific productivity levels per paid employee, indexed for the year 2011 [19,20].

Analyses

All analyses were performed in agreement with the intention-to-treat principle, thus including all participants as randomised. To that end, missing data were imputed. Since substantial dropout had occurred, sensitivity analyses were conducted to gauge the robustness of our findings across different imputation techniques. In the main analysis, missing data were replaced by their most likely value under the expectation maximisation algorithm in SPSS 19. In one sensitivity analysis, all analyses were repeated with last observation carried forward, as implemented in SPSS. In yet another sensitivity analysis, regression imputation as implemented in Stata (version 12.1) was used to impute missing data. As predictor variables we used baseline costs, baseline work functioning, age, gender, partner status and the Karasek factors job demands, control, support from colleagues and superiors [25,26]. Directing the sensitivity analyses towards the various imputation strategies was an *a priori* decision, because it was imperative to ascertain that the research findings did not solely hinge on the chosen imputation technique.

The cost effectiveness analysis was conducted from the societal perspective in which all costs and benefits were included, irrespective of who bears the costs or receives the benefits [9]. Both the incremental costs and incremental effects were used to calculate the incremental cost effectiveness ratio. The incremental cost effectiveness ratio was calculated as $(C_1 - C_0)/(E_1 - E_0)$, where C denotes the average per-participant costs and E is the effect in the experimental and control conditions (subscripted 1 and 0, respectively). The incremental cost effectiveness ratio can be interpreted as the net costs (or savings) per treatment responder.

To handle stochastic uncertainty in the cost and effect data, non-parametric bootstraps were used to simulate 5,000 incremental cost effectiveness ratios. The incremental cost effectiveness ratios were plotted on the cost effectiveness plane to capture the uncertainty in the incremental cost effectiveness estimate (see Figure 2). To be more precise, each simulated incremental cost effectiveness ratio can be plotted on one of the four quadrants of the incremental cost effectiveness plane. In the North East quadrant the intervention produces superior health gains at additional costs relative to the control condition. In the North West quadrant less health is produced for additional costs. Clearly, this is the worst possible outcome, and the intervention is then 'dominated' by the control condition. In the South West quadrant less health is produced, but there are cost savings. Finally, in the South East quadrant the intervention generates superior health gains (relative to the comparator condition) and does so for lower costs. This is the best possible outcome and the intervention is then said to 'dominate' the control condition. It is often seen that a

new intervention falls in the North East quadrant, because better health is obtained for additional costs.

Results

Sample characteristics

Baseline characteristics of the groups are shown in Table 1. There were no differences across the conditions in terms of demographics, baseline costs and work functioning. Therefore we concluded that randomisation had resulted in a balanced trial.

Table 1. Sample characteristics by condition at baseline.

	Control condition (n = 206)		Occupational physician condition (n = 207)		E-mental health condition (n = 204)	
Age, mean (sd)	41.83	(11.3)	42.56	(11.4)	387.5	(12.2)
Female, N (%)	159	(77.2)	170	(82.1)	169	(82.8)
Working hours, mean (sd)	30.98	(6.0)	28.73	(8.0)	31.33	(5.2)
Living with a partner, N (%)	154	(74.8)	153	(73.9)	151	(74)
Born in the Netherlands, N (%)	176	(85.4)	167	(80.7)	174	(85.3)
Work experience, years (sd)	11.3	(10.1)	12.53	(10.4)	10.03	(10)
Turnover intention, N (%)	22	(10.7)	27	(13.0)	25	(12.3)
<i>Baseline costs,^a mean (sd)</i>						
Medication costs	1.06	(6.6)	1.54	(18.7)	1.68	(11.3)
Health care service use	116.97	(229.1)	121.84	(239.5)	211.85	(1,090.4)
Absenteeism	491.62	(1,689.3)	659.92	(2,110.1)	376.98	(856.2)
Presenteeism	1,068.93	(1,862.8)	1,125.04	(2,429.3)	973.65	(1,541.2)
Direct non-medical costs	10.67	(19.87)	10.69	(19.7)	20.1	(104.8)
Work functioning, mean (sd)	14.11	(9.5)	12.56	(9.2)	13.4	(9.2)

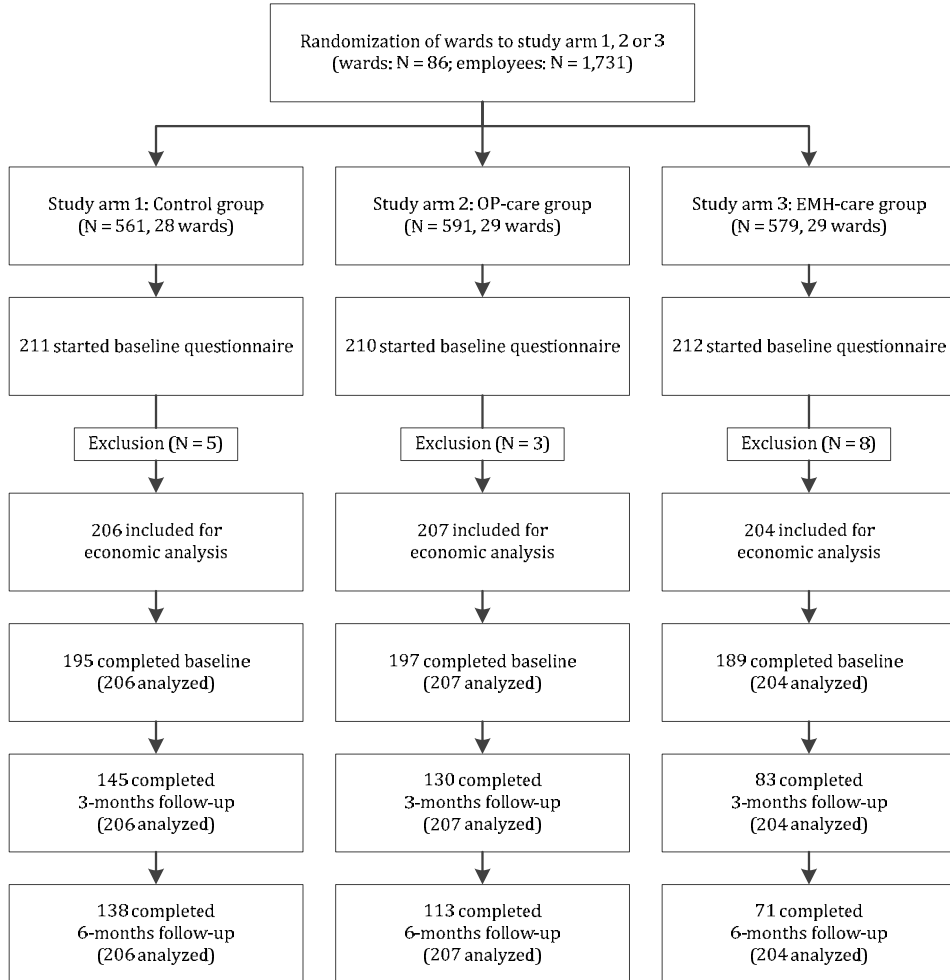
^a In €, time horizon of baseline costs was 3 months.

Missing data and dropout

At baseline, data on impaired work functioning were missing for 11/206 (5%) participants in the control condition, 10/207 (5%) in the occupational physician condition and 15/204 (7%) in the e-mental health condition. At three months follow-up, the dropout rates in the control condition, the occupational physician condition and the e-mental health condition were 61 (30%), 77 (37%) and 121 (59%), respective-

ly. At 6-months follow-up, dropout rates had increased to 68 (33%), 94 (45%), 133 (65%). The flow of the participants through the trial is shown in Figure 1.

Figure 1. Participants' flow through the study



Since loss to follow-up was substantial, we assessed if dropout was selective. A dropout dummy variable (1 = lost, 0 = retained) was computed and regressed on condition, baseline costs, baseline work functioning, age, gender, partner status, and the Karasek factors (job demands, control, support by colleagues and superiors) [25,26]. The analyses indicated that the occupational physician condition was associated with higher dropout than the control condition, and that poorer work functioning at baseline increased the risk of dropout, but having a partner was associated with a smaller likelihood for dropout. When comparing the e-mental health condition to the control condition, it was shown that the e-mental health

condition was associated with greater dropout, as were poorer work functioning and higher job demands at baseline. Again, living together with a partner was associated with a reduced likelihood of dropping out.

Health care service use

The most frequently used health care services among all three conditions were the physiotherapist and GP services. At three months follow-up, consulting the occupational physician increased only in the occupational physician condition, most likely due to the intervention in which these visits were encouraged. Interestingly, the occupational physician condition visits showed a substantial decrease at 6-months follow-up. After six months a reduction in the average volumes of physiotherapist and GP visits is noticeable in all three conditions. Also, use of prescription drugs decreased over time in the three groups.

Supplementary Table 1 lists the average volumes of resource use by treatment group and time (see Appendix 1).

Incremental cost effectiveness

Costs

Table 2 presents the costs of health care uptake and the costs stemming from productivity losses by condition and time point. The most significant costs can be attributed to the productivity losses. In particular, the costs of presenteeism increased in the control condition between baseline and six months follow-up, but decreased in the occupational physician condition and the e-mental health condition.

Incremental costs

Table 3 (upper panel) shows the costs per condition. The average total costs were calculated to be €1,266 per participant in the occupational physician condition and €1,752 in the control condition. The incremental costs were therefore $€1,266 - €1,752 = -€486$ per-participant (negative costs, hence a cost reduction). The incremental costs for the e-mental health condition were associated with a decrease in costs compared to the control condition: $€1,375 - €1,752 = -€377$.

Table 2. Mean per-participant costs by condition and measurement (in €)

	Baseline			3 months			6 months		
	Control (n = 204)	Occupational physician (n = 206)	E-mental health (n = 204)	Control (n = 148)	Occupational physician (n = 132)	E-mental health (n = 89)	Control (n = 140)	Occupational physician (n = 115)	E-mental health (n = 73)
<i>Direct medical costs</i>									
Service use	117	122	212	102	118	102	98	99	100
Medication	1	2	2	1	2	2	1	2	2
<i>Indirect non-medical costs</i>									
Absenteeism	492	660	377	116	348	187	374	234	230
Presenteeism	1,069	1,125	974	1,055	995	995	1,267	916	1,016
<i>Direct non-medical costs</i>	11	11	20	9	11	10	9	11	10

Table 3. Summary statistics for each of the interventions under each imputation technique

	Costs, € ^a	Effect ^b	ICER, € ^c	North East	North West (inferior)	South West	South East (dominant)
Base-case scenario (Expectation maximisation)							
Control	1,752	0.20					
Occupational physician	1,266	0.24	Dominant ^d	2%	1%	21%	75%
E-mental health	1,375	0.16	4,054	1%	8%	76%	16%
Alternative scenario A (Last observation carried forward)							
Control	1,800	0.21					
Occupational physician	1,421	0.27	Dominant ^d	10%	2%	9%	80%
E-mental health	1,388	0.17	6,303	1%	5%	75%	18%
Alternative scenario B (Regression)							
Control	1,681	0.21					
Occupational physician	1,214	0.25	Dominant ^d	1%	0%	17%	81%
E-mental health	1,395	0.16	4,022	1%	8%	74%	16%

^a Cost per treatment responder at 2011 prices | ^b Fraction treatment responders (differences in effect estimates in the text and the Table are due to rounding) | ^c The presented median ICER is 50th percentile of 5,000 bootstrap replications of the ICER | ^d When the ICERs are negative (suggesting cost savings) then they are labelled as 'dominant', representing a situation where the intervention is superior to the control condition from a cost effectiveness perspective.

Incremental effects

Table 3 (upper panel) shows the effects per condition. At follow-up, approximately 24% of the participating nurses met criteria for treatment response in the occupational physician condition compared with 20% in the control condition. In the e-mental health condition, this was about 16%. The incremental effectiveness between the occupational physician condition and the control condition was therefore $0.237 - 0.204 = 0.033$. This was $0.157 - 0.204 = -0.047$ for the e-mental health condition versus the control condition.

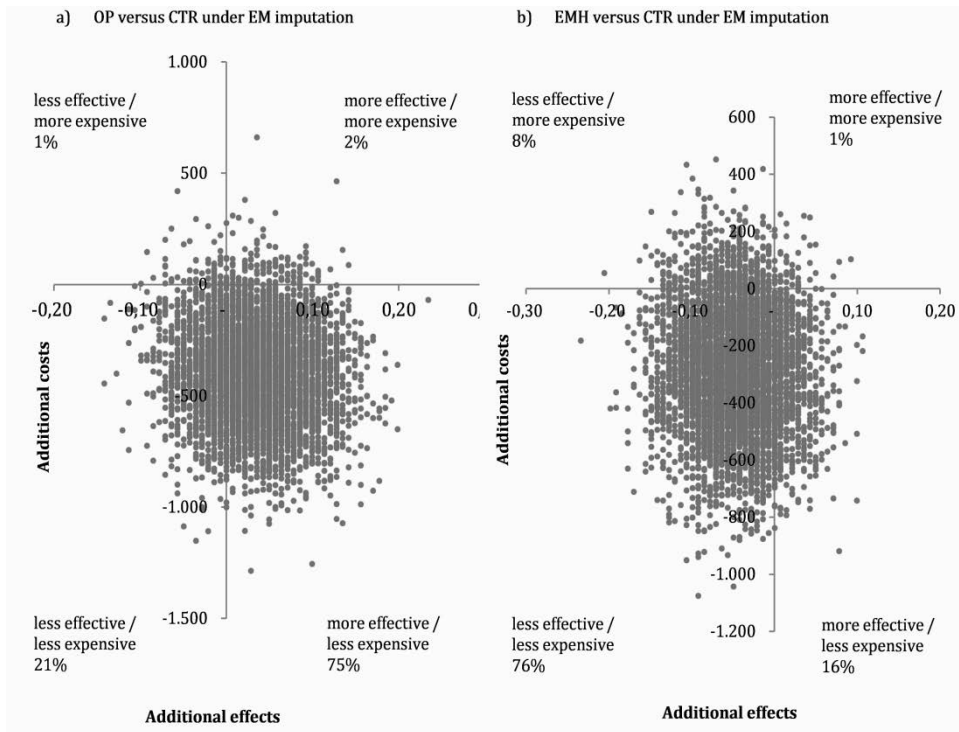
ICER occupational physician condition versus control condition

As noted for the occupational physician condition, the incremental costs were -€486 (savings) and the incremental effect was 0.033. We rely on the median incremental cost effectiveness ratio (ICER) as estimated from the 5,000 non-parametric bootstraps. The median incremental cost effectiveness ratio for the occupational physician condition versus the control condition was estimated as a saving of -€5,049 per treatment responder. Figure 2a shows the scatter of bootstrapped incremental cost effectiveness ratios on the incremental cost effectiveness plane. Of the 5,000 simulated incremental cost effectiveness ratios, 75% fall into the South East-quadrant, indicating that more treatment responses are generated for fewer costs by the occupational physician intervention relative to the control condition. Another 2% of the simulated incremental cost effectiveness ratios fall in the North East-quadrant, indicating a probability of 2% that by applying the intervention an additional treatment response is produced, but at additional costs. The remainder of the stimulated incremental cost effectiveness ratios showed up on the west side of the plane, indicating less effectiveness and less costs (21%), or less effectiveness and more costs (1%). In sum, the occupational physician condition is associated with a 75% probability that the intervention generates better outcomes for less money than the control condition.

ICER e-mental health condition versus control condition

As noted, in the e-mental health condition the incremental costs were -€377 (negative costs, hence a cost saving), but the incremental effect was -0.047 (a small loss in effectiveness) relative to the control condition. The median incremental cost effectiveness ratio could be estimated as €4,054. Figure 2b shows that 76% of the simulated incremental cost effectiveness ratios fall in the South West-quadrant indicating a probability of 76% that by applying the e-mental health condition fewer treatment responses are produced, albeit at less additional costs. Another 16% of the simulated incremental cost effectiveness ratios fell into the South East-quadrant, indicating that more treatment responses are generated for less additional costs by the e-mental health condition relative to the control condition. Finally, 1% indicates more effects at higher costs and 8% indicates less effect at higher costs.

Figure 2. Scatter of simulated incremental cost effectiveness ratios (n = 5,000) on the cost effectiveness plane.



Sensitivity analysis

Sensitivity analyses were conducted by repeating all analyses under two alternative imputation strategies: using last observation carried forward imputation (in SPSS 19) and under regression imputation (in Stata 12.1). Table 3 (lower panels) presents the summary statistics of the sensitivity analyses. As can be seen, the findings that were obtained under expectation maximisation imputation are consistent with the results produced under last observation carried forward and regression imputation. In fact, the results obtained under expectation maximisation imputation fall between last observation carried forward and regression imputation.

Discussion

Main findings

This study was conducted to assess the cost effectiveness of two strategies (the occupational physician condition and the e-mental health condition) to improve work functioning compared to the control condition of screening alone. The proportion of participants that manifested with a reliable change in work functioning was higher, although non-significant, in the occupational physician condition: 23.7% against 20.4% in the control condition. The proportion of improved participants on work functioning in the e-mental health condition on the other hand was lower: 15.7% against 20.4% in the control condition. The average per-responder costs in the three conditions were €1,266 in the occupational physician condition, €1,375 in the e-mental health condition and €1,752 in the control condition.

The median incremental cost effectiveness ratio in the occupational physician condition versus the control condition comparison is dominant (-€5,049), hence a cost-saving per treatment responder. Therefore, the occupational physician condition can be recommended over the control condition, because the occupational physician condition is associated with a greater likelihood that beneficial effects are obtained for fewer costs.

The median incremental cost effectiveness ratio in the e-mental health condition versus the control condition comparison is €4,054 per treatment responder. The e-mental health intervention does not outperform the control condition, because the odds are that nurses derive fewer benefits from this intervention. Sensitivity analyses attested to the robustness of these findings.

Strengths and limitations

Some of the strengths of this study are its randomised design and its relatively large sample size. This is worth mentioning, because the literature identifies several barriers to conducting randomised trials within the work setting such as the difficulties encountered when randomising employees while also trying to avoid the risk of contamination [27-30].

The design with cluster randomisation and pre-randomisation was a strength of this study because randomisation at the ward level reduces the contamination of study groups. Furthermore, pre-randomisation allowed the participants to be blinded with respect to the information related to the other study groups. This minimised contamination effect in the study.

Next, in an attempt to tie the results of this study to previous research, wherein a lack of economic workplace mental health promotion studies is prevailing [8,31], this mental health oriented intervention for nurses can be seen as a welcome addition to the literature strengthening the evidence base.

The findings of this study need to be placed in the context of the study's limitations.

First, the trial suffered from dropout, and our own analysis indicated that dropout had not occurred randomly. Unfortunately we do not have data on reasons for non-responding to the interventions. However, we can speculate that the main reason for drop-out in this specific study population might be feelings of work overload and the inability to find the time to improve their self-management skills. Again, this remains speculative but might have distorted the findings. However, we conducted intention-to-treat analysis using expectation maximisation imputation to handle missing data. In addition, sensitivity analyses were conducted with different imputation techniques and these attested to the robustness of our findings.

Second, the follow-up period used in this trial was short and we do not know what the cost effectiveness of the interventions would look like beyond six months.

Third, the per-participant intervention costs were partly based on assumptions. In particular, the assumption about the number of nurses that would engage in the online screening was important, because the number of participants affects economies-of-scale and ultimately determines the costs of online screening. It should be mentioned that whenever we had to make an assumption, we preferred to err on the conservative side, thus making a conscious choice to steer away from sketching too positive a picture of the cost effectiveness of the interventions. It should also be noted that the costs for screening are low anyway and are therefore unlikely to have a substantial impact on the outcomes overall.

Fourth, the study's results are inevitably conditional on the central clinical endpoint: work functioning. This was an *a priori* choice, but is also a limitation, because in this economic evaluation we did not report on secondary outcomes such as mental wellbeing, changes in symptom level of mental distress and so forth. Had we chosen other outcomes, we would have drawn different conclusions, because the nurses manifested with favourable response on some of these outcomes, even when they did not manifest with treatment response on work functioning. Thus, when we say that the e-mental health intervention is not to be recommended, we say this only with respect to work functioning and our recommendations cannot be generalised towards other outcomes.

Fifth, the measurements of work functioning were based on self-report and this may have biased outcomes. However, it is difficult to say if this might have led to an upward or downward bias. Moreover, we are looking at relative change in work functioning over time and this may have cancelled out a constant bias in participants to exaggerate or diminish the level of their work functioning, while randomisation may have counteracted bias across conditions.

Sixth, it should be noted that all costs are computed for a situation in which the interventions have been fully implemented. Thus the initial investment required to implement the interventions is not part of our study. This was done in agreement with guidelines for economic evaluations [9,20,23], but we recognise that the costs required for implementing the interventions might be interesting in their own right. Estimates of these costs can therefore be obtained from the first author.

Seventh, although we complied with the guidelines for pharmacoeconomic evaluations carried out from the societal perspective whereby direct and indirect costs, inside and outside the healthcare system were included, we acknowledge the possibility of missing costs that might influence the results. Potentially, medical costs which may arise during life-years gained as a result of the treatment are lacking in this study. However, due to the preventive nature of the intervention, it is rather unlikely that these indirect costs within the health care system influence the cost effectiveness results in a substantial manner.

Finally, this study was designed as a pragmatic trial that was conducted in the real-life context of one large academic medical centre in the Netherlands. The strength of this approach is that the trial has a good ecological (external) validity [32]; its weakness is that the outcomes cannot be interpreted as evidence of the interventions' efficacy – only of the interventions' effectiveness under real-life conditions. The hospital in which this study was performed is an academic medical centre. Therefore, the findings are best generalised towards other teaching hospitals, while some caution must be applied when projecting the study's outcomes on hospitals that are not embedded in a university. In this context, it is important to note that the nurses were under constant pressure from their professional obligations and were free to make use of the interventions offered. We see that uptake rates and compliance rates are low, especially regarding the e-mental health interventions. While this may strengthen the level of realism of the trial's outcomes, the outcomes can now not be read as evidence for or against the efficacy of the interventions and are likely to differ from estimates that would have been obtained under tightly controlled conditions.

Recommendations

For improving work functioning in nurses, we recommend implementing the occupational physician condition over the control condition, because the occupational physician condition is associated with better outcomes and cost savings. However, we must be careful recommending for or against implementing the e-mental health condition, because it is associated with a smaller likelihood of producing beneficial effects albeit for lesser costs than the control condition. We also note that had the e-mental health intervention been embedded more rigorously in the work setting, then uptake rates might have looked very different and the e-mental health condition might have yielded more favourable outcomes. At any rate, we recommend that e-health interventions be more fully integrated in the organisation before testing their effectiveness. These recommendations need to be viewed with some caution, because the economic evaluation was conducted with the specific, perhaps somewhat narrow, aim of improving work functioning. Moreover, the outcome was based on self-report, was extended over a brief follow-up period of six months, and was measured in the context of substantial, possibly selective, drop-out.

Acknowledgement

The economic evaluation conducted alongside the Mental Vitality @ Work trial was funded by grant #208010001 from the Netherlands Organization for Health Research and Development (ZonMw) and co-financed by a grant from the Foundation Institute Gak (SIG).

References

1. Campo MA, Weiser S, Koenig KL. Job strain in physical therapists. *Phys Ther* 2009; 89:946-956.
2. Gärtner FR, Nieuwenhuijsen K, Van Dijk FJ, Sluiter JK. The impact of common mental disorders on the work functioning of nurses and allied health professionals: a systematic review. *Int J Nurs Stud* 2010; 47:1047-1061.
3. Magnavita N & Heponiemi T. Violence towards health care workers in a Public Health Care Facility in Italy: a repeated cross-sectional study. *BMC Health Serv Res* 2012; 12:108.
4. Suresh P, Matthews A, Coyne I. Stress and stressors in the clinical environment: a comparative study of fourth-year student nurses and newly qualified general nurses in Ireland. *J Clin Nurs* 2013; 22:770-779.
5. Tayler CM. Subordinate performance appraisal: what nurses really want in their managers. *Can J Nurs Adm* 1992; 5:6-9.
6. Karsh BT, Holden RJ, Alper SJ, Or CK. A human factors engineering paradigm for patient safety: designing to support the performance of the healthcare professional. *Qual Saf Health Care* 2006; 15 Suppl 1:i59-i65.
7. Suzuki K, Ohida T, Kaneita Y et al. Mental health status, shift work, and occupational accidents among hospital nurses in Japan. *J Occup Health* 2004; 46:448-454.
8. Zechmeister I, Kilian R, McDaid D. Is it worth investing in mental health promotion and prevention of mental illness? A systematic review of the evidence from economic evaluations. *BMC Public Health* 2008; 8:20.
9. Drummond M, Brandt A, Luce B, Rovira J. Standardizing methodologies for economic evaluation in health care. Practice, problems, and potential. *Int J Technol Assess Health Care* 1993; 9:26-36.
10. Tompa E, Dolinschi R, De Oliveira C. Practice and potential of economic evaluation of workplace-based interventions for occupational health and safety. *J Occup Rehabil* 2006; 16:375-400.
11. Gärtner FR, Nieuwenhuijsen K, van Dijk FJ, Sluiter JK. Impaired work functioning due to common mental disorders in nurses and allied health professionals: the Nurses Work Functioning Questionnaire. *Int Arch Occup Environ Health* 2012; 85:125-138.
12. ILO. Technical and ethical guidelines for workers' health surveillance. *Scand J Work Environ Health* 1998; 24:158.
13. Koh D & Aw TC. Surveillance in occupational health. *Occup Environ Med* 2003; 60:705-710.
14. Gärtner FR, Ketelaar SM, Smeets O et al. The Mental Vitality @ Work study: design of a randomized controlled trial on the effect of a workers' health surveillance mental module for nurses and allied health professionals. *BMC Public Health* 2011; 11:290.
15. Schellings R, Kessels AG, Sturmans F. Pre-randomisation in study designs: getting past the taboo [in Dutch: Prerandomisatie bij wetenschappelijk onderzoek: een taboe doorbroken]. *Ned Tijdschr Geneesk* 2008; 152:2053-6.
16. Schellings R, Kessels AG, ter Riet G et al. Indications and requirements for the use of pre-randomization. *J Clin Epidemiol* 2009; 62:393-399.
17. Gärtner FR, Nieuwenhuijsen K, van Dijk FJ, Sluiter JK. Psychometric properties of the Nurses Work Functioning Questionnaire (NWFQ). *PLoS One* 2011; 6:e26565.
18. Gärtner FR, Nieuwenhuijsen K, van Dijk FJ, Sluiter JK. Interpretability of change in the Nurses Work Functioning Questionnaire: minimal important change and smallest detectable change. *J Clin Epidemiol* 2012; 65:1337-1347.
19. CBS (Centraal Bureau voor de Statistiek), 2012. [http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=71311NED&D1=0,2,4,6&D2=0-1,61,70,87,108,137,145,172,176,221-222,230,255,1&D3=\(1-34\)-1&HD=081020-1310&HDR=T&STB=G1,G2](http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=71311NED&D1=0,2,4,6&D2=0-1,61,70,87,108,137,145,172,176,221-222,230,255,1&D3=(1-34)-1&HD=081020-1310&HDR=T&STB=G1,G2). Accessed September 2012.
20. Hakkaart L, Tan SS, Bouwmans, CAM. Manual for cost research. Methods and standards costs for economic evaluations in healthcare [in Dutch: Handleiding voor kostenonderzoek. Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg]. Rotterdam, the Netherlands: Instituut voor Medical Technology Assessment, Erasmus Universiteit Rotterdam, 2010.

CHAPTER 5

21. Healthcare institute Netherlands. Pharmacotherapeutic compass [in Dutch: College van Zorgverzekeringen. Farmacotheurapeutisch kompas]. [Cited September 2012]; Available from: <http://www.fk.cvz.nl/>.
22. Koopmanschap MA. PRODISQ: a modular questionnaire on productivity and disease for economic evaluation studies. *Expert Rev Pharmacoecon Outcomes Res* 2005; 5:23-28.
23. Krol M, Papenburg J, Koopmanschap M, Brouwer W. 2011. Do productivity costs matter?: the impact of including productivity costs on the incremental costs of interventions targeted at depressive disorders. *Pharmacoeconomics* 2005; 29:601-619.
24. Weisbrod BA. The Valuation of Human Capital. *J Polit Econ* 1961; 69:164-170.
25. Karasek R, Baker D, Marxer F et al. Job decision latitude, job demands, and cardiovascular disease: a prospective study of Swedish men. *Am J Public Health* 1981; 71:694-705.
26. Karasek R, Choi B, Ostergren P-O et al. Testing two methods to create comparable scale scores between the Job Content Questionnaire (JCQ) and JCQ-like questionnaires in the European JACE study. *Int J Behav Med* 2007; 14:189-201.
27. Chapman R & Combs S. Collaboration in the Emergency Department: an innovative approach. *Accid Emerg Nurs* 2005; 13:63-69.
28. Chapman R, Duggan R, Combs S. Leading Change and Advancing Health by Enhancing Nurses' and Midwives' Knowledge, Ability and Confidence to Conduct Research through a Clinical Scholar Program in Western Australia. *ISRN Nurs* 2011; 2011:245417.
29. Chau JP, Lopez V, Thompson DR. A survey of Hong Kong nurses' perceptions of barriers to and facilitators of research utilization. *Res Nurs Health* 2008; 31:640-649.
30. Kajermo KN, Unden M, Gardulf A et al. Predictors of nurses' perceptions of barriers to research utilization. *J Nurs Manag* 2008; 16:305-314.
31. Mihalopoulos C, Vos T, Pirkis J, Carter R. The economic analysis of prevention in mental health programs. *Annu Rev Clin Psychol* 2011; 7:169-201.
32. Ramsey S, Willke R, Briggs A et al. Good research practices for cost effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. *Value Health* 2005; 8:521-533.

Appendix 1. Supplementary Table 1: Average volumes of cost items by condition over time

	Baseline			3 months			6 months		
	CTR (n = 204)	OP (n = 206)	EMH (n = 204)	CTR (n = 148)	OP (n = 132)	EMH (n = 89)	CTR (n = 140)	OP (n = 115)	EMH (n = 73)
Services, N (%)	11 (5)	10 (5)	15 (7)	11 (7)	8 (6)	2 (2)	10 (7)	3 (3)	3 (4)
Psychologist	3 (1)	1 (<1)	3 (1)	1 (1)	1 (1)	1 (1)	2 (1)	2 (2)	0 (0)
Psychiatrist	70 (34)	73 (35)	78 (38)	45 (30)	41 (31)	28 (31)	46 (33)	29 (25)	20 (27)
GP	14 (7)	21 (10)	13 (6)	8 (5)	25 (19)	6 (7)	8 (6)	4 (3)	4 (5)
OP	24 (12)	30 (15)	34 (17)	17 (11)	20 (15)	15 (17)	16 (11)	12 (10)	11 (15)
Physiotherapist	12 (6)	10 (5)	4 (2)	9 (6)	6 (5)	5 (6)	8 (6)	4 (3)	4 (5)
Coach	1 (<1)	2 (1)	4 (2)	5 (3)	1 (1)	1 (1)	2 (1)	2 (2)	0 (0)
Company Social Worker	2 (1)	1 (<1)	3 (1)	3 (2)	1 (1)	0 (0)	3 (2)	2 (2)	0 (0)
Social Worker	2 (1)	1 (<1)	0 (0)	1 (1)	2 (2)	1 (1)	1 (1)	2 (2)	2 (3)
Spiritual counsellor	10 (5)	10 (5)	18 (9)	7 (5)	6 (5)	4 (4)	9 (6)	9 (8)	4 (5)
Alternative treatment	15 (7)	10 (5)	13 (6)	8 (5)	6 (5)	7 (8)	5 (4)	7 (6)	3 (4)
Medication ^a									

^a All types of medication are included.

CHAPTER

6

Protecting and promoting mental health of nurses in the hospital setting: Is it cost effective from an employer's perspective?

Cindy Noben
Filip Smit
Karen Nieuwenhuijsen
Sarah Ketelaar
Fania Gärtner
Judith Sluiter
Silvia Evers

Accepted for publication: International Journal of Occupational Medicine and Environmental Health

Abstract

Objectives: Nurses are at elevated risk of burnout, anxiety and depressive disorders, and may then become less productive. This begs the question if a preventive intervention in the work setting might be cost-saving from a business perspective.

Methods: A cost benefit analysis was conducted to evaluate the balance between the costs of a preventive intervention among nurses at elevated risk of mental health complaints and the cost offsets stemming from improved productivity. This evaluation was conducted alongside a cluster-randomized trial in a Dutch academic hospital. The control condition consisted of screening without feedback and unrestricted access to usual care (N=206). In the experimental condition screen-positive nurses received personalized feedback and referral to the occupational physician (N=207).

Results: Subtracting intervention costs from the cost offsets due to reduced absenteeism and presenteeism resulted in net-savings of €244 per nurse when only absenteeism is regarded, and €651 when presenteeism is also taken into account. This corresponds with a return-on-investment of €5 up to €11 for every Euro invested.

Conclusions: Within half a year, the costs of offering the preventive intervention were more than recouped. Offering the preventive intervention represents a favourable business case as seen from the employer's perspective.

Keywords: prevention, nurses, occupational health, mental disorders, cost benefit, work functioning.

Accepted for publication: International Journal of Occupational Medicine and Environmental Health

Introduction

Some nurses are at elevated risk for stress and mental health problems due to high job demands and a lack of autonomy [1, 2]. Poor mental health is undesirable in its own right, but it may also have financial implications for the employer [3, 4] via absenteeism, presenteeism (reduced at-work job performance) and staff turnover [5, 6]. From a business point of view it might therefore be of value to protect and promote mental health in nurses and maintain the quality of their work.

Periodic screening could be useful to detect early signs of mental health complaints and personalized feedback could encourage help-seeking among nurses. A Workers' Health Surveillance (WHS) instrument was developed for this purpose. The WHS is a preventive strategy that aims at the early detection of negative health effects and work functioning problems and includes personalized feedback. The WHS is followed up by referral to the occupational physician (OP) for screen-positive nurses in need of intervention. This three-tiered intervention aims to detect mental health problems in the earliest stages and prevent further deterioration of these problems. In so doing, the intervention may also enhance job performance [7, 8].

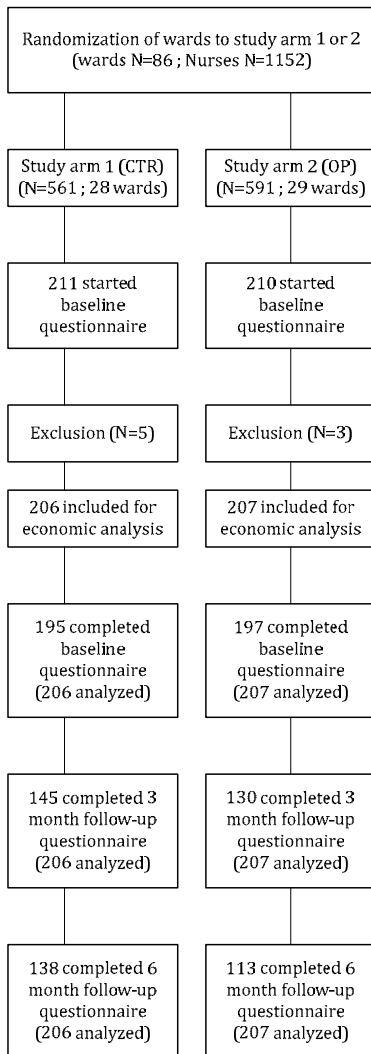
Elsewhere, we published a cost effectiveness analysis of the intervention from the societal perspective [9]. That study took account of the costs of health care uptake, pharmacy use and nurses' out-of-pocket expenses for travelling to health care services. The outcome of interest was treatment response. It was concluded that screening, feedback and OP care led to improved work functioning and these were associated with a 75% likelihood of lower costs than a 'do nothing' scenario as seen from a societal perspective. However, an employer is likely to look at a different set of financial parameters to inform decisions about implementing an intervention in the work setting. This paper adopts the employer's perspective and assesses whether providing screening followed by personalized feedback and referral to the OP represents a viable business case. In contrast to the aforementioned cost effectiveness analysis, we now look at the costs that are incurred by the employer of offering the preventive intervention. These costs are then compared with the benefits (expressed in Euro, €) that are, again, relevant from the employer's perspective, such as the costs differences stemming from reduced absenteeism and improved productivity while at work. In short, this paper is conducted as a cost benefit analysis to address the question if the benefits outweigh the costs. If this were the case, then the net-benefits would be suggestive of a favourable business case that may persuade employers to implement the preventive WHS intervention in the work setting.

Methods

Study design

The study was conducted in an academic medical centre in the Netherlands as a pragmatic cluster randomized controlled trial with randomization at the level of hospital wards. A cost benefit analysis was conducted from the employer's perspective to see if there is a business case for investing in the employees' mental health and work functioning. All costs were calculated in Euro for the reference year 2011 using the consumer price index from Statistics Netherlands [10]. For the current cost benefit analysis we compared two conditions: (1) the OP condition (screening, feedback followed by referral to the OP for the screen positives), versus (2) the control (CTR) condition (screening without feedback and without referral to the OP). Within the hospital, 29 wards (with 207 consenting nurses) were randomized to the OP condition and 28 wards (with 206 consenting nurses) to the CTR condition. Data were collected at baseline and after three and six months (henceforth t_0 , t_1 and t_2). Both costs and benefits were computed over a six-month time horizon, corresponding to the follow-up period of the study. We excluded healthcare costs (other than those attributable to the intervention) and nurses' out-of-pocket costs for obtaining health care, because they were deemed not to be relevant from the employer's perspective. Costs and benefits were not discounted because the follow-up time did not exceed one year. A medical ethics committee approved the study. Figure 1 presents the flow of participants through the trial. More information regarding the design of the Mental Vitality @ Work study can be obtained elsewhere [11].

Figure 1. Flow of participants



Intervention and control conditions

All participants were screened for work functioning impairments and six types of mental health complaints: distress, work-related fatigue, risky drinking, depression, anxiety, and post-traumatic stress disorder. Nurses in the CTR condition filled out the screening questionnaire and no further steps were taken. In the OP condition, screening was followed by personalized feedback and screen-positive nurses received an invitation to visit the occupational physician. The subsequent OP consultation was structured according to a seven-step protocol, with the focus on identify-

ing impairments in work functioning and providing advice on how to improve well-being and work functioning. The seven-step protocol included the following: i) discussing expectations; ii) discussing screening results and characteristics of work functioning and mental health complaints; iii) discussing possible causes in the private, work, and health condition and consequences for work functioning; iv) identifying the problem and offering rationale; v) giving advice on how to tackle the health complaints, how to improve work functioning, how to prevent consequences of impaired work functioning, and how to communicate with the supervisor about work functioning and mental health; vi) discussing possible follow-up or referral to other care providers; and vii) summarising the consultation. All participating OPs received three-hour training in using the protocol [12].

Computation of intervention costs

The costs of offering the intervention included (1) the costs of operating the web-based screening and feedback module, (2) the costs for periodically upgrading the module and (3) the costs of hosting the module on a server (including maintenance costs). These costs amounted to €4 per user (calculations can be obtained from the first author). Furthermore, the per-participant costs for consulting the OP (€73) and the costs for the OP-assistant for scheduling the nurses' visits to the OP (€3) are also included. To these we added the costs of training the OPs in using the preventive consultation protocol (€50 per OP visit). Thus, a nurse who engaged in screening, received feedback and made a single visit to the OP would cost €130. However, it needs to be borne in mind that some screen-positive nurses did not visit their OP, while others made a single visit or multiple visits.

Computation of the benefits

As seen from the employer's perspective, the benefits from the intervention are related to increased productivity levels due to reduced absenteeism and presenteeism. Changes in productivity were valued in monetary terms using the human capital method. This method assesses the loss of productivity by multiplying the self-reported number of working days lost due to absenteeism multiplied by the average gross gender and age specific wages per paid employee. The wage estimation was according to the Dutch guideline for health economic evaluation and can be found in Table 1 [13-15].

Table 1. Productivity by gender and age groups (in €)

Age groups	Men	Women
15 - 19	9.88	8.97
20 - 24	18.18	17.59
25 - 29	24.77	24.19
30 - 35	30.37	28.20
35 - 40	34.85	29.96
40 - 45	37.25	29.76
45 - 50	39.25	29.61
50 - 55	40.00	29.96
55 - 60	40.33	30.21
60 - 65	40.07	29.36

The work days lost due to diminished productivity were based on the self-reported number of work days when the nurse did not feel well while at work over the past 6 months, weighted by an inefficiency score derived from the Productivity and Disease Questionnaire (PRODISQ)[16]. This was done on a 10-point rating scale, ranging from 0 to 1, with 0 meaning not inefficient and 1 completely inefficient. The number of work days lost due to inefficiency was then multiplied with gender and age-specific wages indexed for the year 2011 [10, 15]. Finally, the benefits were computed by comparing the pre-intervention costs (at t_0) with those post intervention (at t_2). This yielded a pre-post cost difference in each condition and these could then be compared across the conditions.

Cost benefit analysis

All analyses were performed in agreement with the intention-to-treat principle, thus including all participants as randomized. In the main analysis, missing data were replaced by their most likely value under the expectation maximization (EM) algorithm in SPSS 19.

The incremental costs, C , were the intervention costs of the OP condition minus the intervention costs of the CTR condition. The incremental benefits, B , were computed as the cost savings due to reduced productivity losses (owing to pre-post changes in both absenteeism and presenteeism) in the OP condition minus the cost savings in the CTR condition. Net-benefits were computed as $B-C$, the cost-to-benefit ratio as C/B and the return on investment (ROI) as B/C .

The net benefits, cost-to-benefit ratio and return on investment were analysed in Stata (version 12.1) using non-parametric bootstrap techniques. The analyses took into account that observations were clustered, as nurses were 'nested' in different wards at the hospital. Therefore robust sample errors were obtained using the first-order Taylor series linearization within each of the 1,000 bootstrap steps. This procedure was conducted on the dataset that was imputed using EM.

Sensitivity analysis

Sensitivity analyses were conducted to assess the robustness of our findings by making less optimistic assumptions about the benefits. In this context it is of note that the benefits due to reduced presenteeism were computed by multiplying the inefficiency score by the number of days at work with diminished work productivity. However, it may be assumed that presenteeism may not impact on productivity levels when the diminished productivity is compensated for during normal working hours by the nurse or by colleagues [17]. If that is true, then we may have produced an overly optimistic estimate of the benefits. Thus to test the robustness of our findings, we recomputed the cost benefit ratio by reducing the benefits by 10%, 20% and 30%, and by omitting the cost offsets of reduced presenteeism altogether.

Results

Sample characteristics

Baseline characteristics of the groups are shown in Table 2. Both groups were quite similar regarding demographic and occupational characteristics. The majority of the participants were female nurses, born in the Netherlands who lived together with a partner. On average the participants were aged 42 years and had more than 10 years of work experience. We concluded that randomization had resulted in a balanced trial.

Table 2. Baseline characteristics of the sample

	CTR group (n=206)	OP group (n=207)
Age, mean (SD)	41.83 (11.3)	42.56 (11.4)
Female, N (%)	159 (77.2)	170 (82.1)
Function, N (%)		
Nurse	146 (70.9)	124 (59.9)
Surgical nurse	9 (4.4)	14 (6.8)
Nurse practitioner	23 (11.2)	14 (6.8)
Allied health professional	21 (10.2)	31 (15)
Anaesthetic nurse	0	13 (6.3)
Other	7 (3.4)	11 (5.3)
Working hours, mean (SD)	30.98 (6)	28.73 (8.1)
Living with a partner, N (%)	154 (74.8)	153 (73.9)
Born in the Netherlands, N (%)	176 (85.4)	167 (80.7)
Work experience, years (SD)	11.3 (10.1)	12.53 (10.4)
Turnover intention, N (%)	22 (10.7)	27 (13)

CTR=Control | OP= Occupational Physician

Cost benefit analysis

Table 3 presents the per-nurse intervention costs and benefits in the OP and the control condition as well as the net-benefits.

The mean per-nurse intervention costs were €89 in the OP condition and €25 in the CTR condition. The cost difference between the conditions was therefore €64 (95%CI=52~76), which was statistically significant (robust bootstrapped SE=6.03; z=10.5; p<0.001).

Cost reductions due to greater productivity were €715 (95% CI= 226~1,203) in the OP condition relative to those in the CTR condition. These cost savings were statistically significant (robust bootstrapped SE=249; z=2.87; p=0.004) and in favor of the OP condition.

Subtracting per-participant intervention costs from the per-participant cost offsets due to reduced absenteeism and presenteeism resulted in net-savings of €651per nurse. The net-benefits were statistically significant (95%CI=167~1,135; SE=247.13; z=2.63; p=0.008) and in favour of the OP condition. Benefits stemming from reduced presenteeism are hard-to-quantify, by excluding these benefits and focusing on net-benefits when only absenteeism is regarded resulted in net-benefits of €244 per nurse in favour of the OP condition, still representing a favourable business case as seen from the employer's perspective.

Table 3. Intervention costs, benefits and net-benefits (per nurse, in €)

		OP group	CTR group	DIFF (OP-CTR)
Costs	Screening	4	4	0
	Added costs OP training	50	0	50
	OP care t0	13	6	7
	OP care t1	17	7	10
	OP care t2	5	7	-3
	<i>Total intervention costs</i>	<i>89</i>	<i>25</i>	<i>64</i>
Benefits	Absenteeism t0	660	492	
	Absenteeism t2	234	374	
	<i>Averted absenteeism cost</i>	<i>426</i>	<i>118</i>	<i>308</i>
	Presenteeism t0	1,125	1,069	
	Presenteeism t2	916	1,267	
	<i>Averted presenteeism cost</i>	<i>209</i>	<i>-198</i>	<i>407</i>
<i>Total benefits</i>	<i>635</i>	<i>-80</i>	<i>715</i>	
Net benefits	<i>Net- benefits (presenteeism included)</i>	<i>546</i>	<i>-105</i>	<i>651</i>
	<i>Net- benefits (presenteeism excluded)</i>	<i>337</i>	<i>93</i>	<i>244</i>

CTR=Control | OP= Occupational Physician | DIFF (OP-CTR) = Difference (Occupational Physician group minus Control group)

Base-case and sensitivity analyses

In the base-case analysis the return on investment for the OP condition was €7 per invested Euro and -3 Euros per invested Euro for the CTR condition (i.e. negative

benefits, thus higher costs). The return on investment was (715/64=) €11 per invested Euro (see Table 4).

Various sensitivity analyses were performed to attest the robustness of the findings and the results are summarized in Table 4. The results of the sensitivity analyses attest to the robustness of the main analysis. The incremental costs of offering the intervention are more than compensated for by productivity gains. Even when productivity gains in the OP-condition are lowered by 30% the net benefit per employee is still €461 after six months and the return-on-investment is still a substantial €8 per invested Euro. When ignoring the hard-to-quantify benefits stemming from reduced presenteeism, there still is a return on investment of almost €5.

Table 4. Results of base-case and sensitivity analyses in €

		OP	CTR	DIFF (OP-CTR)	Net-benefits (B-C)	Cost benefit ratio (C/B)	Return on investment (B/C)
Base- Case	Costs (C)	89	25	64			
	Benefits (B)	635	-80	715	651	0.09	11
Sensitivity Analyses	-10% OP Benefit	571	-80	651	587	0.11	10
	-20% OP Benefit	508	-80	588	524	0.11	9
	-30% OP Benefit	444	-80	525	461	0.12	8
	-Presenteeism	426	118	308	244	0.21	5

CTR=Control | OP= Occupational Physician | DIFF (OP-CTR) = Difference (Occupational Physician group minus Control group) | B = Benefits | C = Costs

Discussion

Main findings

The primary aim of the study was to conduct a cost benefit analysis from the employer's perspective by considering the balance of the costs of a preventive intervention and the cost offsets stemming from improved productivity. Net-benefits were €651 per nurse and were statistically significant at $p=0.008$. In other words, the pay-out is €11 per Euro invested. It is worth noting that the cost offsets occur within six months post intervention, thus representing a favourable business case from an employer's perspective where the initial investments are more than recouped within a short period of time. The main conclusion that offering the intervention offers good value for money from the employer's perspective remains intact when confining the analysis to the changes in absenteeism and ignoring any benefits due to reduced presenteeism.

Placing the results in the wider context of the literature

There are only a few studies evaluating the costs and benefits of mental health promotion and prevention in the workplace. Knapp, McDaid and Parsonage have assessed the economic impact of mental health and well-being improvements associated with various programs based on a limited range of studies using economic modelling [18]. Arends and colleagues evaluated a return-to-work intervention and focused on the prevention of recurrent sickness absence and helping workers to stay at work. The authors demonstrated the 12-months effectiveness of a problem-solving intervention for reducing recurrent sickness absence in workers with common mental disorders [19]. Iijima and colleagues conducted a cost benefit analysis of mental health prevention programs within Japanese workplaces. They concluded that the majority of companies gained a net benefit from the mental health prevention programs [20]. Another Japanese study concluded that a participatory work environment improvement program and individual-oriented stress management programs showed better cost benefits, suggesting primary prevention programs for mental health at the workplace economically advantage employers [21]. Although the results of our study unite with the results of previously conducted studies, which took place during and after the sickness absence period, or within another cultural, ethical and geographical setting, our study might be seen as a welcome addition to a limited evidence-base.

Strengths and limitations

Several strengths of this study need to be mentioned. First, this study was a trial-based economic evaluation and is therefore firmly rooted in empirical data. Second, the study was conducted as a pragmatic trial, thus enhancing the ecological validity of the results [22]. Third, we reviewed both the cost reductions due to less absenteeism and reduced presenteeism. Although including presenteeism is standard practice in economic evaluations, it is worth mentioning that owing to their lower visibility these costs are not so apparent to an employer and are often overlooked [13, 23].

The results need to be placed in the context of the study's limitations. First, the trial data were affected by drop-out and this may have distorted the outcomes. That said, we conducted an intention-to-treat analysis by imputing missing observations under the EM algorithm. In our health economic evaluation of the same data we demonstrated that the results after EM imputation are very similar to those obtained under alternative imputation strategies such as regression imputation and last-observation-carried-forward imputation. Nonetheless, drop-out rates were substantial and may have biased our outcomes.

Second, all outcomes were based on self-report. However, it is hard to see how presenteeism can be measured without resorting to self-assessments of diminished productivity. Unfortunately, the validity of self-reported presenteeism has not often

been researched [24-28]. It is precisely for this reason that we conducted sensitivity analyses to gauge the robustness of the study's outcomes when less optimistic assumptions are being made about the benefits in the OP condition.

Third, some potential impacts of the intervention were not assessed, such as the costs of staff turnover and the spill-over effects of absenteeism by one nurse on the workload of her colleagues. As to staff turnover, the data indicate that turnover intention was reduced from 27 nurses at t_0 down to ten nurses at t_2 in the OP condition. In the CTR condition these were 22 and 14, respectively. These data suggest that the OP intervention may have additional favourable effects on staff turnover, but these were not quantifiable in terms of actual changes in staff turnover. As a consequence we can now only speculate that the cost benefits that we reported represent lower bounds of the true cost benefits.

Fourth, it should be noted that all intervention costs are computed for. Thus the initial investments required for developing and implementing the interventions were part of our study. Although this might contradict with guidelines for economic evaluations [29], were one would solely account for intervention costs when fully implementing the intervention, we recognized that the costs required for developing and implementing the interventions were interesting in their own right and therefore included in the total intervention costs.

Fifth, there are two main approaches to costing productivity losses; the human capital approach and the friction cost approach. However, both approaches produce similar results in the short term, as is the case in our study with a follow-up after six months. We therefore expect that choosing one approach or the other is unlikely to have a major impact on our conclusions. From a business case point of view the relatively short follow-up time is not a limitation, because it is good to see that the costs of offering the intervention are recouped within such short time span. Nevertheless, as yet we do not have data to inform us about the longer-term outcomes.

Conclusions

A return-on-investment of €11 within six months represents a very appealing business case, and wider implementation of the intervention can be recommended. However, and as noted above, the time horizon of this study is limited in that we only look at costs and cost reductions over a six month period. Therefore, we do not know if effects are maintained over time. In all likelihood the intervention's impact will need to be sustained by periodic repetition of the intervention, e.g. a screening plus personalized feedback plus referral to the OP for screen-positives every 12 or 24 months.

References

1. Tayler CM. Subordinate performance appraisal: what nurses really want in their managers. *Can J Nurs Adm.* 1992 Sep-Oct;5(3):6-9.
2. Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA.* 2003 Jun 18;289(23):3095-3105.
3. Goetzel RZ, Hawkins K, Ozminkowski RJ, Wang S. The health and productivity cost burden of the "top 10" physical and mental health conditions affecting six large U.S. employers in 1999. *J Occup Environ Med.* 2003 Jan;45(1):5-14.
4. Kessler RC, Greenberg PE, Mickelson KD, Meneades LM, Wang PS. The effects of chronic medical conditions on work loss and work cutback. *J Occup Environ Med.* 2001 Mar;43(3):218-225.
5. Lerner D, Amick BC, Lee JC, et al. Relationship of employee-reported work limitations to work productivity. *Med Care.* 2003 May;41(5):649-659.
6. Adler DA, McLaughlin TJ, Rogers WH, et al. Job performance deficits due to depression. *Am J Psychiatry.* 2006 Sep;163(9):1569-1576.
7. Gärtner FR, Nieuwenhuijsen K, van Dijk FJ, Sluiter JK. The impact of common mental disorders on the work functioning of nurses and allied health professionals: a systematic review. *Int J Nurs Stud.* 2010 Aug;47(8):1047-1061.
8. Gärtner FR, Nieuwenhuijsen K, van Dijk FJ, Sluiter JK. Impaired work functioning due to common mental disorders in nurses and allied health professionals: the Nurses Work Functioning Questionnaire. *Int Arch Occup Environ Health.* 2012 Feb;85(2):125-138.
9. Noben CY, Smit F, Nieuwenhuijsen K, et al. Comparative cost effectiveness of two interventions to promote work functioning by targeting mental health complaints among nurses: pragmatic cluster randomised trial. *International Journal of Nursing Studies.* 2014.
10. Centraal Bureau voor de Statistiek 2012 [cited 2012 September 2012]. Available from: [http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=71311NED&D1=0,2,4,6&D2=0-1,61,70,87,108,137,145,172,176,221-222,230,255,1&D3=\(1-34\)-1&HD=081020-1310&HDR=T&STB=G1,G2](http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=71311NED&D1=0,2,4,6&D2=0-1,61,70,87,108,137,145,172,176,221-222,230,255,1&D3=(1-34)-1&HD=081020-1310&HDR=T&STB=G1,G2).
11. Gartner FR, Ketelaar SM, Smeets O, et al. The Mental Vitality @ Work study: design of a randomized controlled trial on the effect of a workers' health surveillance mental module for nurses and allied health professionals. *BMC Public Health.* 2011 May 10;11.
12. Ketelaar SM, Nieuwenhuijsen K, Gartner FR, et al. Mental Vitality @ Work: The effectiveness of a mental module for workers' health surveillance for nurses and allied health professionals, comparing two approaches in a cluster-randomised controlled trial. *Int Arch Occup Environ Health.* 2013. Epub 2013 Jul 28.
13. Krol M, Papenburg J, Koopmanschap M, Brouwer W. Do productivity costs matter?: the impact of including productivity costs on the incremental costs of interventions targeted at depressive disorders. *Pharmacoeconomics.* 2011 Jul;29(7):601-619.
14. Weisbrod BA. The Valuation of Human Capital *The Journal of Political Economy* 1961;69(5):164-170.
15. Hakkaart L, Tan SS, Bouwmans, CAM. Manual for cost research. Methods and standards costs for economic evaluations in healthcare [in Dutch: Handleiding voor kostenonderzoek. Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg]. Rotterdam, the Netherlands: Instituut voor Medical Technology Assessment, Erasmus Universiteit Rotterdam, 2010.
16. Koopmanschap MA. PRODISQ: a modular questionnaire on productivity and disease for economic evaluation studies. *Expert Rev Pharmacoecon Outcomes Res.* 2005 Feb;5(1):23-28.
17. Krol M, Brouwer W, Rutten F. Productivity Costs in Economic Evaluations: Past, Present, Future. *Pharmacoeconomics.* 2013 Apr 26(7):537-549.
18. Knapp M, McDaid D, Parsonage M. Mental Health Promotion and Prevention: The Economic Case. London: Department of Health 2011.
19. Arends I, van der Klink JJ, van Rhenen W, de Boer MR, Bultmann U. Prevention of recurrent sickness absence in workers with common mental disorders: results of a cluster-randomised controlled trial. *Occup Environ Med.* 2014 Jan;71(1):21-29.

CHAPTER 6

20. Iijima S, Yokoyama K, Kitamura F, Fukuda T, Inaba R. Cost benefit analysis of comprehensive mental health prevention programs in Japanese workplaces: a pilot study. *Ind Health*. 2013;51(6):627-633.
21. Yoshimura K, Kawakami N, Tsusumi A, et al. Cost benefit analysis of primary prevention programs for mental health at the workplace in Japan. *Journal of occupational health*. 2013 Feb 26;55(1):11-24.
22. Ramsey S, Willke R, Briggs A, et al. Good research practices for cost effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. *Value Health*. 2005 Sep-Oct;8(5):521-533.
23. Krol M, Brouwer W, Sendi P. Productivity costs in health-state valuations : does explicit instruction matter? *Pharmacoeconomics*. 2006;24(4):401-414.
24. Beaton DE, Tang K, Gignac MA, et al. Reliability, validity, and responsiveness of five at-work productivity measures in patients with rheumatoid arthritis or osteoarthritis. *Arthritis Care Res* 2010 Jan 15;62(1):28-37.
25. Prasad M, Wahlqvist P, Shikiar R, Shih YC. A review of self-report instruments measuring health-related work productivity: a patient-reported outcomes perspective. *Pharmacoeconomics*. 2004;22(4):225-244.
26. Sanderson K, Tilse E, Nicholson J, Oldenburg B, Graves N. Which presenteeism measures are more sensitive to depression and anxiety? *J Affect Disord*. 2007 Aug;101(1-3):65-74.
27. Short ME, Goetzel RZ, Pei X, et al. How accurate are self-reports? Analysis of self-reported health care utilization and absence when compared with administrative data. *J Occup Environ Med*. 2009 Jul;51(7):786-796.
28. Zhang W, Bansback N, Anis AH. Measuring and valuing productivity loss due to poor health: A critical review. *Soc Sci Med*. 2011 Jan;72(2):185-192.
29. Drummond M, Brandt A, Luce B, Rovira J. Standardizing methodologies for economic evaluation in health care. Practice, problems, and potential. *Int J Technol Assess Health Care*. 1993 Winter;9(1):26-36.

CHAPTER

7

A pilot economic evaluation to improve a web-based employability intervention for employees

Cindy Noben
Silvia Evers
Joost van Genabeek
Frans Nijhuis
Angelique de Rijk

Submitted

Abstract

Objectives: To perform a pilot cost effectiveness study for a web-based employability intervention aiming at improving intervention content and study design.

Methods: Uptake rate analysis for the intervention elements, cost effectiveness, cost utility and subgroup analyses were conducted to identify potential content-related intervention improvements. Differences in work ability and quality adjusted life years and overall contribution of resource items to the total costs were assessed. These were used to guide study design improvements.

Results: 63 participants were a-select allocated to either the intervention (n=29) or the control (n=34) group. Uptake was $\leq 70\%$. Cost effectiveness and cost utility analyses resulted in negative effects although higher total costs. Incremental effects were marginal (work ability -0.51; QALY -0.01).

Conclusions: The intervention needs to be improved with respect to targeting and intensity; outcome measures selected and collection of cost data.

Keywords: pilot study; feasibility; cost effectiveness; web-based intervention; occupational health; QALY

Submitted

Introduction

Since 2000 research has established that sickness absence results in considerable negative effects on work for employees, employers, and society as a whole. Absence from work due to sickness absence constitutes a large economic burden, especially for those on long-term sickness absence [1, 2]. Employees with health complaints, chronic illness or disability often worry about the sustainability of their employment. Due to their health complaints, problems can arise at work, primarily with respect to either obtaining work or remaining at work. Franche *et al.* and Amick *et al.* in 2002 and 2004 respectively, provided evidence regarding the readiness for return to work and predictors for successful work functioning [3, 4]. Both studies found significantly improved self-efficacy among supportive organizations that acknowledged the importance of psychosocial management [4] and the interplay between care, workplace and insurance [3]. Additionally, the importance of improving an employee's self-efficacy to maintain successful work-role functioning and improve work ability has been established [3-6]. More recently, Hoefsmit *et al.* (2013) found that in order to improve work ability, cooperation between employee and employer is essential [7]. The study showed that in an attempt to achieve stakeholder cooperation, employees and employers need to: 'be aware of their mutual dependence'; 'trust each other'; and 'have knowledge about the legislation'. Likewise, employees with higher levels of self-efficacy are better able to stay at work despite their health problems and may even return to work faster after reporting sick [7]. Furthermore, a recently published study showed that return-to-work self-efficacy was a significant predictor of reintegration to work among very long-term sick-listed employees [8]. Thus, improving self-efficacy may empower an employee to handle difficulties in relation to his or her health problems [9].

In recent years, the internet is being used more often to locate information on a broad range of topics, including work-health related issues. Previous research has investigated the reach, compliance and perceived effectiveness of interactive websites or e-therapies aimed at empowering work-disabled employees. However, these interventions occurred either prior to the disability assessment undertaken by an insurance physicians [10, 11] or focused solely on information provision and helping patients to formulate questions to ask their physicians [12, 13]. Because of the increasing number of internet users, a growing number of interactive interventions are delivered online.

To increase self-perceived work ability (i.e. self-efficacy) among employees who experience problems in relation to their employability and work ability, a web-based employability intervention was developed [14]. A previously conducted study among unemployed people assessed the effectiveness of the web-based employability intervention [15]. These analyzes showed that the intervention had no effect on the longing to return to work. Furthermore, no intervention effect was found for the period within which the participant thinks to be able to (return to)

work. The effects of the web-based employability intervention were contrary to what was expected since participants in the control group were more often at work during the post-test instead of the participants in the intervention group. When conducting separate analyses for the different intervention components, only the knowledge website itself showed a small significant ($\beta=0.18$ at 0.05 significance level) improvement on the degree of self-efficacy. Neither the other intervention components (i.e. personal advice, the forum, or the feedback session) nor the complete web-based employability intervention, were able to detect an effect on degree of self-control, empowerment, or return to work [15].

Therefore, a critical look at the current content and design of the web-based employability intervention is needed. However, detailed information on the (cost) effectiveness of web-based employability interventions is often lacking or conflicting. Therefore, key analytical decisions regarding the most important data to collect and sample sizes are calculated based on smaller pilot trials [16]. A pilot study can provide information about the feasibility to conduct (cost-) effectiveness trials, which can be costly and time-consuming. This saves valuable resources being unnecessarily spent on large scale studies that may not be feasible. Typically when conducting a pilot study, the intervention is assumed to be effective and the content is therefore stated. This pilot study highlights how potential improvements in the study design can be detected and improved. This is done by critically assessing the content of the intervention with respect to the uptake of the various intervention elements, and by studying the effectiveness in diverse employee groups.

Therefore, the purpose of this study was to:

- (i) detect whether based on the results of the pilot economic evaluation, the web-based employability intervention content can be improved to improve widespread implementation;
- (ii) detect whether based on the results of the pilot economic evaluation of the web-based intervention, the study design could be improved to guide future research.

Materials and methods

The data collection for this pilot study was conducted in the Netherlands between June 2013 and March 2014 with a follow-up period of three months. The web-based employability intervention was compared with regular trade union support among participants with work-related health problems. Members of the trade union can call or use the instant messaging service via the website to ask questions about work, health and disease. They are then attended to a union consultant who answers the questions or refers the caller to a disciplined expert. Participants were approached via the participating employees' trade union in the public sector. Members of this union predominately worked for privatized government agencies or in the care and welfare sector. Eligible participants were adults aged between 18-65

years. Participants either had concerns regarding their ability to continue their work responsibilities, were at a crossroad of calling in sick for a longer period of time, or were absent for more than six weeks and were feeling unable to return to work in the near future. Those who were absent from work for more than two years were excluded from the study. Advertisements in the trade unions' newsletters targeted the participant group and based on individual's own intrinsic motivation, could be recruited into the research. After the completion of the first questionnaire participants were randomly divided into either the control group or the intervention group. Participants and researchers were not blinded for either the treatment allocation or intervention uptake. In the control group, a reward of €10 was provided to those who completed both questionnaires. In the intervention group, participant received free usage of the web-based employability intervention for three months.

Intervention and Control Condition

Participants from the control group could contact the participating trade union for regular support (e.g. information on website, telephone contact, etc.) regarding their work- and health related concerns. The control group did not have access to the web-based employability intervention. This was only accessed by the intervention group.

The web-based employability intervention focused on improving the self-efficacy of employees by encouraging and motivating them to take control of their own work context and enhance their work ability. The web-based employability intervention focused on teaching employees skills to return to work and/or stay at work. In particular, the training allowed the participants to ask job and health-related questions, as well as guiding communication with relevant stakeholders (e.g. supervisors, colleagues, occupational physicians, medical practitioners, labor experts, job coaches). Specifically, the intervention aimed to 'empower' employees to access regular, curative and occupational health care services during their illness, recovery and work reintegration. The four core elements of the intervention were: (i) a knowledge based web-site of current information focusing on work ability in relation to health; (ii) an assessment of the work ability of respondents (Work Ability Index-test), followed by a subsequent telephone interview providing respondents with feedback on their results; (iii) personal and technical expert advice via phone and email; and (iv) an online interactive forum on work and health for sharing worker's experiences, questions and answers with content experts. Using the intervention elements was voluntary. However, a subscription to the knowledge site was needed (<http://www.wijzermetwelder.nl/>).

Aim I: Evaluating the content of the web-based employability intervention

The first aim of this pilot economic evaluation was to consider whom to include and how to target suitable participants for the web-based employability intervention in the future. The intervention aimed to provide an easily accessible method for interested people to ask questions about disability, health and work. Furthermore, the intervention was suitable for everybody, regardless of their health or disability. Hence, a number of features of the intervention were studied. First, the uptake rates for each of the four core elements of the intervention were examined. This identified potential improvements for one or more of the core intervention elements. Recommendations regarding compliance and uptake could then be made. Second, a cost effectiveness analysis (CEA) and a cost utility analysis (CUA) were conducted by accounting for all costs, irrespective of who pays or bears them. Sub-group analyses were conducted based on age (under and above 50 years), sex (males versus females) and working hours (more than 20 hours per week), to provide information on how to better target the intervention.

Aim II: Evaluating the study design of the web-based employability intervention

The second aim of this study was to improve the study design of web-based employability interventions by using the results of the pilot economic evaluation to carefully select outcomes and cost-categories that matter both to the employee and society as whole. Differences in work ability and quality adjusted life years (QALY) among both groups were assessed. Work ability was measured via “The Work Ability Index” (WAI), by considering both the physical and mental demands of the work. The index helps to identify those that need support in these areas [17]. A point improvement on the WAI denoted a substantial improvement and hence, demonstrated a gained effect (i.e. better work ability). Differences in Quality Adjusted Life Years (QALYs) were calculated using the EuroQol-5D-5L [18, 19]. A crosswalk to the 3-level value sets was used to calculate the utility weights based on EuroQol scores using Dutch tariffs [20, 21]. Utilities at baseline and after three months follow-up were used to compute QALYs, by means of the area under the curve method [22]. In order to assess the importance of resource items in their overall contribution to the total costs, the frequency of consumption and their unit costs were assessed. Questionnaires measuring healthcare resources were provided to the respondents at baseline and three months after the intervention started. Healthcare costs (including primary, secondary, alternative care, hospitalization, and medication), costs to patient-and family (including travel- and parking, and informal care costs), and productivity costs (including costs related to absenteeism and presenteeism) were estimated and indexed for the reference year 2013 [23-26]. Costs of the web-based employability intervention included those related to implementing and operating the intervention.

Statistical analysis

Missing data per respondent were replaced by mean imputation using the mean of the series of the appropriate outcome. The analyses required complete cases responses for both baseline and three month follow-up measures. Therefore, respondents who were lost to follow-up were excluded from the analyses. Baseline characteristics of both groups were compared using t-tests of variances, or by the non-parametric equivalent. That is, Kruskal-Wallis for analysis of continuous variables or χ^2 test for categorical comparisons of proportions. The costs of both groups were compared by the non-parametric bootstrapping method using 95% confidence intervals in percentiles. By bootstrapping, samples of the same size as the original data were drawn with replacements from the observed data. An incremental approach was used to compare the employability intervention with regular trade union support. The incremental costs per 'improved' respondent on the WAI scale and per QALY gained (i.e. the incremental cost effectiveness ratios (ICER)) were graphically plotted on one of the four quadrants of the cost effectiveness planes. For this, non-parametric bootstrap re-sampling techniques were therefore used. In the North West quadrant if fewer effects were produced for additional costs, then the intervention was considered "dominated" by the control condition. In the South East quadrant, if the intervention generated superior effects (relative to the comparator condition) with fewer costs; the intervention was considered "dominating" the control condition. Detailed calculations relevant for the economic evaluation are found in the *Technical Appendix*.

Results

Participants

Between June 2013 and March 2014, 70 potential participants were eligible for inclusion, provided their informed consent, and were allocated to either the web-based employability intervention (n=34) or the control (usual care) group (n=36). In total, 63 participants (29 in the intervention group; 34 in the control group) completed both the baseline and the three month follow-up questionnaires and were included in the analyses. The demographic characteristics of the participants are shown in Table 1. At baseline, more females were allocated to the intervention group. A significant difference was found for mean health scores with participants in the intervention group rating their health lower than the control group. Aside from gender and self-rated health, no significant differences between the two groups with respect to their demographic characteristics were found.

Table 1. Baseline characteristics (N=63)

	Intervention (n=29)	Group	Control (n=34)	Group	<i>p</i> value
Age, mean (sd)	51 (8)		53 (7.9)		0.4(a)
Female, N (%)	22 (76)		18 (53)		0.06(b)
Education [§] , N (%)					
Low	2 (7)		3 (9)		0.8(b)
Intermediate	10 (34)		14 (41)		
High	17 (59)		17 (50)		
Working hours, mean (sd)	26.5 (8.9)		29.3 (8.5)		0.2(a)
Living situation, N (%)					
Living alone	5 (17)		8 (24)		0.4(b)
Living alone with children	-		2 (6)		
Living together with a partner	14 (48)		17(50)		
Living together with partner and children	10 (35)		7 (20)		
Job demands, N (%)					
Physical	2 (7)		1 (3)		0.8(b)
Psychological	14 (48)		18 (53)		
Both physical and psychological	13 (45)		15 (44)		
Function, N (%)					
Industrial sector	-		2 (6)		0.14(b)
Agricultural sector	1 (3)		-		
Administrative sector	3 (10)		10 (29)		
Service sector	1 (3)		-		
Healthcare sector	15 (52)		9 (27)		
Research & Development	2 (7)		3 (9)		
Other	7 (24)		10 (29)		
Social benefits, N (%)					
WIA / WGA / WAJONG	3 (10)		3 (9)		0.6(b)
Other	2 (7)		1 (3)		
None	24 (83)		30 (88)		
Permanent employment, N (%)	27 (93)		33 (97)		0.4(b)
Absenteeism days, mean (sd)	5.47 (8.6)		3.21 (6.9)		0.15(c)
Self-rated health [‡] , mean (sd)	67 (12.3)		75 (17.2)		0.015(c)
Costs, mean in € (sd)					
Intervention	56.51 (22)		-		-
Medication	350.86 (926)		197.86 (679)		0.465(c)
Healthcare	644.91 (858)		918.50 (3052)		0.187(c)
Patient and Family	68.39 (120)		48.53 (61.93)		0.341(c)
Absenteeism	847.05 (1439)		543.87 (1195)		0.243(c)
Presenteeism	241.87 (296)		311.81 (283)		0.189(c)

[§]Low=preschool or primary school; Intermediate=lower- or upper secondary; High=tertiary school, university or postgraduate

[‡]Self-reported health score ranging 0 (worst) to 100(best)

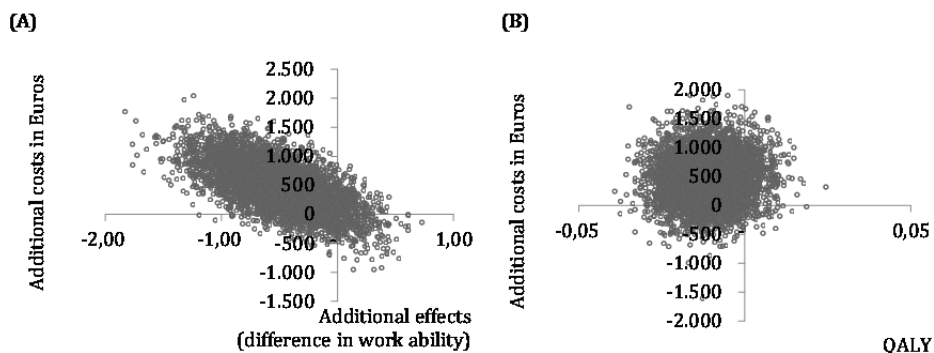
(a) ANOVA; (b) χ^2 test; (c) Kruskal-Wallis Test

Improving the content and targeting of the web-based employability intervention

The uptake rates for each of the four core elements of the intervention were low. Out of the 29 participants in the intervention group, 50% (n=15) logged on to the knowledge site, approximately 70% (n=20) of the respondents participated in a telephone interview after completing the WAI-test to discuss their results, 16 requested personal and technical expert advice, and only one person used the online interactive forum.

Marginally lower score on the WAI (-0.51) and higher total costs (€483.8) for the intervention group (when compared to the control group) resulted in a negative ICER (Δ total costs/ Δ total effects) (see also Table 2). The majority of the incremental cost-effect pairs were located in the northwest quadrant indicating fewer effects, albeit higher costs, for the web-based employability intervention (Figure 1). The three-month ICER resulted in a large negative incremental cost utility ratio surrounded with large uncertainty which can be seen in the wide distribution of ICERs (Figure 1). 82% of the bootstrapped cost utility pairs were situated in the northwest quadrant, representing less effectiveness (i.e. less quality of life gains) and more costs for the intervention group compared to the control group.

Figure 1. Cost effectiveness planes for (A) cost effectiveness analysis, and (B) cost utility analysis



The results of the subgroup analyses are presented in the lower level in Table 2. The CEA for the subgroups of respondents aged above 50 years and female genders resulted in a minimal increase in work ability (0.03 and 0.05 points improvement on the WAI respectively), associated with higher costs compared with the respondents in the usual care condition. In order for the intervention to be effective, a point improvement on the WAI was required. However, the results of both subgroup analyses should be interpreted with caution as one cannot state that the web-based employability intervention was more effective although it had higher associated

CHAPTER 7

costs. Furthermore, based on the distribution of ICERs on the CE plane, these estimates are not robust. The subgroup of employees working more than 20 hours per week resulted in lower work ability in the web-based employability intervention group compared to the control group. There were also higher total costs with a negative ICER of -457.69. However, it is important to note that the ICER does not fall within the confidence interval (Table 2).

Table 2. Mean bootstrapped cost and effect differences between subgroups of the intervention- and control group including incremental cost effectiveness ratios, 95% CIs and cost effectiveness plane distributions

Analysis	Sample Size		ΔC Euro	ΔE WAI QALY	ICER (95% CI)	Distribution of CE plane [‡]			
	Intervention	Control				%NE	%SE (dominant)	%SW	%NW (inferior)
<i>Total group</i>									
CEA ^v	29	34	483.8	-0.51	-948.37 (-1,164;335)	5	4	7	84
CUA [‡]	29	34	483.8	-0.01	-45,045.13 (-51,078;1,526)	6	3	9	82
<i>CEA Subgroup</i>									
Aged under 50	12	12	345.94	-0.67	-518.91 (-496;-3,133)	3	7	16	75
Aged above 50	17	22	9.74	0.03	280 (-88;2,318)	13	41	8	38
Females	22	18	58.01	0.05	1,276.16 (5,450;10,667)	19	35	10	36
Males	7	16	1,147.36	-1.37	-839.90 (-970;151)	4	4	2	91
Working hours >20	21	28	261.54	-0.57	-457.69 (1,354;4,588)	1	7	22	70

^vIn the CEA ΔE is the mean difference in work ability (WAI score), ΔC is the mean difference in societal costs

[‡]In the CUA ΔE is the mean difference in QALYs, ΔC is the mean difference in societal costs

[‡] NE refers to north-east quadrant of the cost effectiveness plane, which indicates the intervention is more effective and more costly than the control. SE refers to the south-east quadrant of the cost effectiveness plane, which indicates the intervention is more effective and less costly than the control. SW refers to south-west quadrant of the cost effectiveness plane, which indicates the intervention is less effective and less costly than the control. NW refers to north-west quadrant of the cost effectiveness plane, which indicates the intervention is less effective and more costly than the control.

Improving the study design of the web-based employability intervention

Mean incremental differences of -0.51 point improvement on work-ability and mean incremental differences of -0.01 QALY were unable to demonstrate an effect of the web-based intervention (Table 2).

Participants who received the intervention consulted their physiotherapist and medical specialist more often than participants in the control condition. They used more alternative care and had longer stays in the hospital. Number of visits to the general practitioner and the medical specialist were almost equal in both groups. The number of hours of informal care was larger in the control group when compared to the intervention group. Table 3 presents the uncorrected costs of different healthcare utilization costs after three months. As costs were not distributed normally (Table 1), aggregated costs per category (healthcare costs, patient and family costs, costs in other sectors, and societal costs) were bootstrapped. The highest total societal costs were seen in the intervention group (€1,707) vs. (€1,223) in the control group. The relative contribution of patient-and family out-of-pocket costs was smallest; whilst productivity costs contributed for almost 50% of the total societal costs in both groups (Table 3).

Table 3. Cost prices, mean total costs (SD) and mean bootstrapped costs (95% CI) during the 3-month follow-up (n=63)

Cost type	Cost price per unit(€)**	Mean costs (sd)		Mean bootstrapped costs (95% CI)‡	
		Intervention Group	Control Group	Intervention Group	Control Group
Healthcare costs		399.7 (386.1)	365.9 (566.99)	399 (270;543)	366 (194;574)
General practitioner	30.5	37.1 (63.3)	40.59 (66)	36 (17;62)	40 (22;65)
Physiotherapist	33.7	165.1 (180.77)	102.1 (180.72)	165 (91;250)	102 (49;165)
Occupational physician	38.1	16.5 (40)	26.39 (48.23)	17 (5;33)	26 (12;43)
Medical specialist	34.3	27 (40.32)	25 (47.21)	27 (14;43)	25 (11;42)
Psychologist	94.3	31.5 (80.2)	91 (216.04)	30 (4;64)	90 (27;169)
Hospital admission	549.7	19.3 (102.03)	0.27 (1.55)	19 (0;57)	0
Hospital care day parts	273.2	19.3(101.41)	8.43 (46.84)	19 (0;57)	8 (0;25)
Alternative care	67.1	75.6 (170)	21.1 (60.78)	75 (24;144)	21 (4;44)
Home care	71.6	8.5 (40.11)	54.47 (248.85)	9 (0;24)	51 (0;143)
Medication costs	Variable*	280.3 (846.36)	147.9 (299.62)	281 (84;627)	148 (69;262)
Patient and family costs	Variable†	43.5 (45.4)	46.9 (64.28)	43 (29;61)	47 (28;69)
Costs in other sectors		890.1 (1,334.35)	662.2 (1,028.5)	890 (457;1,411)	661 (364;1,040)
Absenteeism	Age & gender specific wages§	724.4 (1,330.65)	464.5 (1,064.96)	723 (293;1,250)	462 (150;853)
Presenteeism	Age & gender specific wages§	165.7 (71.98)	197.7 (89.22)	166(141;192)	198 (168;226)
Total social costs		1,706.8 (1675.1)†	1,223 (1,539.97)	1,707 (1153;2327)	1,223 (758;1767)

‡ The upper and lower confidence limits are the 2.5th and 97.5th percentile based on 10,000 bootstrap replications
** Standard costs for the Netherlands, 2013 index
‡Dutch Pharmacotherapeutic Compass
†Including informal care costs and travel and parking costs
§Dutch guidelines for costing studies
¶ Deviation from the sum due to the inclusion of per-participant intervention costs

Conclusion

This research demonstrates a number of benefits for undertaking a pilot economic evaluation to improve both the intervention and the study design. Based on the results of this pilot, an attempt is made to bridge the gap between the application of research and a real-life setting. Furthermore, this pilot economic evaluation is unique as it also closely examines the intervention content and provides recommendations for further development.

The first aim of the study was to critically assess the intervention content and detect whether, based on the results of the pilot economic evaluation, the content of the web-based employability intervention should be modified to improve widespread implementation. The findings of the pilot economic evaluation do not indicate a widespread implementation of the web-based employability intervention in practice. Both content related improvements, as well as several practical implications are required. First, compliance and uptake rates need to be increased before implementing a full trial. To do so all four elements of the intervention may need to be improved. In particular, the content of the online interactive forum should be further investigated, as uptake of this component was very low. These findings, to improve all four components, are supported by the qualitative study conducted to assess whether the intervention complied to the needs of the interviewed intervention users [14]. Results have shown that the web-based employability intervention was insufficient to connect the specific needs and wishes of the users. The interviewees stated that the intervention provided via the website and via contacts with experts was too general for their specific situation. Interviewees for the qualitative evaluation had relatively prolonged health problems and often very specific questions about work and health [14]. Additionally, the voluntary usage of all four elements should be reconsidered and methods to stimulate ongoing involvement with web-based employability interventions need to be found. Furthermore, difficulties were experienced when recruiting employees to participate in the study, and in increasing participation- and intervention uptake rates. As a result, future interventions may need to specifically consider the current workplace culture, company specific policies regarding absenteeism and sick-leave, and the employee's own perceived health problems. Careful evaluations incorporating both quantitative- and qualitative methods will be needed to document how the web-based employability intervention can be operationalized and potentially adapted for use in other settings. Secondly, the results of this pilot economic evaluation showed that the web-based employability intervention was not associated with superior clinical or economic impact when compared to regular trade union support. The main analyses indicated that the intervention is being dominated by the control condition (i.e. more costs and fewer effects as the ICER is located in the inferior quadrant of the cost effectiveness plane). The results for this heterogeneous population were

reasonably robust. However, there was a tendency females and participants aged 50 or older might be more sensitive for an intervention effect.

The second aim of the study was to evaluate whether, based on the results of the pilot economic evaluation, the study design should be improved. To improve the study design implementing a web-based employability intervention, outcomes and cost-categories that matter both to the employee and society need to be carefully selected. In this pilot economic evaluation, no significant effects in favour of the intervention were found. However in this study, only the intervention itself was expected to yield differences in outcomes. The true effect of the intervention might have been influenced by participants using information other than the information derived via the web-based employability intervention. It may be disputed whether the most appropriate outcome was selected for this pilot economic evaluation, and whether it was able to measure the true effect of the intervention. Additionally, the intervention might be used in combination with other resources (e.g. social networking, other websites, etc.). The added value of the intervention should be made clear as the effects might also depend on the personal characteristics of the individual user (e.g. healthy or unhealthy behaviour) and on the interaction with alternative information and support resources. Expert opinions or other forms of qualitative research on the appropriate size of quantitatively important distinction between baseline and follow-up might be helpful. Secondly, given the time and effort required to collect cost effectiveness data, this pilot indicates that the collection of patient- and family costs might be omitted based on their relative contribution to the total costs. Productivity costs on the other hand appear to be essential in the economic evaluation of web-based employability.

Discussion

The study results should be placed in the light of other research conducted within the field of e-health and web-based interventions. A recently conducted review summarized the effectiveness (empowerment related outcomes) of several web-based interventions. They concluded that when small positive effects on empowerment were found, the duration and intensity of the interventions were much higher than the web-based employability intervention described here [9]. For clinical problems and the work ability assessment of employees, a long-term time span might have been more appropriate to capture all consequences of the intervention under study. The limited follow-up period (3 months) of this pilot might also mask the visibility of a true effect (e.g. increase in work ability over a longer period of time). However, another study with a longer follow-up period (6 months) was also unable to detect improvements in work functioning [27]. They indicated that the reasons for poor implementation of the e-health intervention included the intervention not fully being integrated into the organisation before testing its effectiveness. These are similar finding to our pilot study which had low uptake rates.

In addition, although randomised controlled trials (RCT) are considered the best method to examine the effectiveness of an intervention or treatment, this pilot questions the feasibility of conducting an RCT in a pragmatic setting. The use of web-based employability interventions within a non-disease specific population raises the issue of generalizability of the results versus the applicability of the results in a disease specific population. As mentioned before, no significant effects were found. On the other hand, because of the inclusion of many types of work-related disabilities, the generalizability of the results of the RCT is upturned. Considering the challenges of using RCT methods to study the effectiveness of the intervention, observational methods or research using qualitative methods, are needed to explore the barriers to conducting RCTs which compare web-based employability interventions to regular trade union support. Consequently, a large RCT appears to be very expensive and infeasible.

Implications for future research

Before conducting a full-scale study, several key elements need to be improved. At this time, the web-based employability intervention was not associated with an economically relevant impact over regular trade union support for a heterogeneous population with work- and disability related concerns. However, based on the knowledge gained from this pilot economic evaluation, the administration of data collection as well as deciding upon cost categories and outcome measures can be used to direct future research. An intervention should attempt to be easily accessible for interested people to ask questions about disability, health and work. The web-based tool might also benefit from better 'targeting' given the heterogeneous population and from revising and enlarging its intensity.

Acknowledgement

This work was financially supported by Foundation Institute Gak. We would like to thank Marjet Veldhuis, Jenny Huijs, Catrien Funke and Kerst Zwart for the data collection, data entry and supplying the raw data. We would like to acknowledge Susan Peters, Senior Research Coordinator at Brisbane Hand and Upper Limb Research Institute, Australia for assistance with English editing.

References

1. Allebeck P, Mastekaasa A. Swedish Council on Technology Assessment in Health Care (SBU). Chapter 5. Risk factors for sick leave - general studies. *Scand J Public Health Suppl.* 2004;63:49-108.
2. Waddell G. *The Back Pain Revolution* Edinburgh: Churchill Livingstone 2000.
3. Franche RL, Krause N. Readiness for return to work following injury or illness: conceptualizing the interpersonal impact of health care, workplace, and insurance factors. *J Occup Rehabil.* 2002 Dec;12(4):233-56.
4. Amick BC, Habeck RV, Ossmann J, Fossel AH, Keller R, Katz JN. Predictors of successful work role functioning after carpal tunnel release surgery. *J Occup Environ Med.* 2004 May;46(5):490-500.
5. Arnstein P. The mediation of disability by self efficacy in different samples of chronic pain patients. *Disabil Rehabil.* 2000 Nov 20;22(17):794-801.
6. Shaw WS, Huang YH. Concerns and expectations about returning to work with low back pain: identifying themes from focus groups and semi-structured interviews. *Disabil Rehabil.* 2005 Nov 15;27(21):1269-81.
7. Hoefsmit N, de Rijk A, Houkes I. Work resumption at the price of distrust: a qualitative study on return to work legislation in the Netherlands. *BMC Public Health.* 2013;13:153.
8. Volker D, Zijlstra-Vlasveld MC, Brouwers EP, van Lomwel AG, van der Feltz-Cornelis CM. Return-to-Work Self-Efficacy and Actual Return to Work Among Long-Term Sick-Listed Employees. *J Occup Rehabil.* 2014 Oct 30.
9. Samoocha D, Bruinvels DJ, Elbers NA, Anema JR, van der Beek AJ. Effectiveness of web-based interventions on patient empowerment: a systematic review and meta-analysis. *Journal of medical Internet research.* 2010;12(2):e23.
10. Samoocha D, Snels IA, Bruinvels DJ, Anema JR, Kowalczyk W, van der Beek AJ. Process evaluation of a web-based intervention aimed at empowerment of disability benefit claimants. *BMC Med Inform Decis Mak.* 2011;11:10.
11. Samoocha D, Snels IA, Bruinvels DJ, Anema JR, van der Beek AJ. Effectiveness of an interactive website aimed at empowerment of disability benefit claimants: results of a pragmatic randomized controlled trial. *J Occup Rehabil.* 2011 Sep;21(3):410-20.
12. Haywood K, Marshall S, Fitzpatrick R. Patient participation in the consultation process: a structured review of intervention strategies. *Patient Educ Couns.* 2006 Oct;63(1-2):12-23.
13. Kinnersley P, Edwards A, Hood K, Ryan R, Prout H, Cadbury N, et al. Interventions before consultations to help patients address their information needs by encouraging question asking: systematic review. *BMJ.* 2008;337:a485.
14. van Genabeek J, Noben C, Rooijackers B, Roozeboom MB, Huijs J, Hengel KO, et al. Does knowledge offers the solution? Research report of an investigation into the usage of a web-based employability intervention among members of a trade union [in Dutch: Biedt inzicht uitzicht? Verslag van een onderzoek naar het gebruik van de interventie 'Wijzer met Welder' door een groep leden van CNV Publieke Zaak.] Report. Hoofddorp: TNO, 2014 R10770 TNO 2014.
15. van Genabeek J, Fermin BMF, Huijs JJM, Huson A, Giesen FBM, Hengel KMO. Employability among the unemployed. Report of a study towards the effectiveness of a web-based employability intervention among unemployed workers. [in Dutch: Wijzer voor vangnetters. Verslag van een onderzoek naar de inzet van de interventie 'Wijzer met Welder' bij vangnetters]. Report. Hoofddorp: TNO, 2014 27 January 2014. Report No.: TNO 2014 R10155.
16. Drummond M, Coyle D. The role of pilot studies in the economic evaluation of health technologies. *Int J Technol Assess Health Care.* 1998 Summer;14(3):405-18.
17. Tuomi K IJ, Jahkola A, Katajarinne L, Tulkki A. *Work Ability Index*. 2nd revised edn. Helsinki: Finnish Institute of Occupational Health. 1998.
18. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res.* 2011 Dec;20(10):1727-36.
19. EQ-5D User Guide. Rotterdam Erasmus Univeriteit Rotterdam, centrum voor gezondheidsbeleid en recht, 1995.

CHAPTER 7

20. Lamers LM, McDonnell J, Stalmeier PF, Krabbe PF, Busschbach JJ. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Econ.* 2006 Oct;15(10):1121-32.
21. EuroQol. [cited 2014]. Available from: <http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-value-sets.html>.
22. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost effectiveness analysis: the importance of controlling for baseline utility. *Health Econ.* 2005 May;14(5):487-96.
23. Oostenbrink JB, Koopmanschap MA, Rutten FF. Standardisation of costs: the Dutch Manual for Costing in economic evaluations. *Pharmacoeconomics.* 2002;20(7):443-54.
24. Healthcare institute Netherlands. Pharmacotherapeutic compass [in Dutch: College van Zorgverzekeringen. Farmacotherapeutisch kompas]. [Cited September 2012]; Available from: <http://www.fk.cvz.nl/>.
25. Drug costs in the Netherlands [cited 2014]. Available from: <http://www.medicijnkosten.nl/>.
26. Hakkaart L, Tan SS, Bouwmans, CAM. Manual for cost research. Methods and standards costs for economic evaluations in healthcare [in Dutch: Handleiding voor kostenonderzoek. Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg]. Rotterdam, the Netherlands: Instituut voor Medical Technology Assessment, Erasmus Universiteit Rotterdam, 2010.
27. Noben C, Smit F, Nieuwenhuijsen K, Ketelaar S, Gartner F, Boon B, et al. Comparative cost effectiveness of two interventions to promote work functioning by targeting mental health complaints among nurses: Pragmatic cluster randomised trial. *Int J Nurs Stud.* 2014 Oct;51(10):1321-31.

Technical Appendix

Cost calculation

Questionnaires measuring the use of resources with a three month recall period were provided to the respondents at baseline and three months after commencement of the intervention to identify healthcare, patient-and family and productivity costs (indexed to the reference year 2013).

Health care costs comprised visits for primary, secondary, alternative care, hospitalization, and medication. The Dutch Manual for Costing was used to value health care utilization. For prescribed medication, prices were calculated by first determining the standard daily dose as reported in the Dutch Pharmacotherapeutic Compass and then multiplied the price per dose by the number of days.

Patient and family costs included out-of-pocket costs (i.e. travel- and parking costs and informal care costs). Travel- and parking costs incurred while making use of health services are valued by computing the costs per kilometer to a health service, adding parking costs, for the health care resource used. The reference price for informal care was based on the hourly minimum wage of a housekeeper at 40 hours per week from the age of 23 years.

Productivity costs (i.e. costs in other sectors) comprised losses in productivity from paid work (absenteeism) and costs due to reduced productivity while at work (presenteeism). Sick leave hours were derived from the total duration of sick leave days in the previous three months by converting them into work-hours equivalents based on the Dutch average of 1,540 work hours per year. The human capital theory was used to value productivity losses by multiplying the sick-leave hours by the gender and age-specific productivity levels per paid employee.

Costs of the web-based employability intervention were calculated by using a bottom-up costing approach and included those related to implementing and operating the employability intervention (Table 1). In depth cost calculation methods can be obtained by contacting the corresponding author.

Table 1. Costs (all in Euro) of the web-based employability intervention

Intervention Component	Units	Unit Prices	Annual costs (n=336)	Participant costs
<i>(i) Operational costs website 'Wijzer met Welder'</i>				
Maintenance ICT	52	45	2,340	6.96
Actualisation content	52	65	3,380	10.06
Content support	260	45	11,700	34.82
Hosting, telephone, etc.	‡	NA	3,500	10.42
Handling – administration	52	35	1,820	5.42
Customer management	16	80	1,280	3.81
Marge	‡	NA	1,500	4.46
Total costs per participant per year			25,520	75.95
Total costs per participant - 3month follow-up				18.99
<i>(ii) Utilization costs WAI-test</i>				
Price arrangement (invoice) - licence: NA		NA	1,337	3.98
WAI-test including questionnaires, updates and protocol				
Total costs per participant per year				3.98
Total costs per participant - 3month follow-up				1.00
<i>(iii) Feedback coach</i>				
Facilitating	10 minutes	47.7	NA	7.95
Preparation	10 minutes	47.7	NA	7.95
Conversation	30 minutes	47.7	NA	23.85
Registration	1 minute	47.7	NA	0.795
Total costs per participant per uptake				40.55
<i>(iv) Personal Advice</i>				
Mailing or advisory telephone line	15 minutes	47.7	NA	11.925
Registration	5 minutes	47.7	NA	3.975
Total costs per participant per uptake				15.9
<i>(v) Forum</i>				
Moderator	5 hours per week	95	24,700	73.51
Total costs per participant per uptake				73.51

‡ average estimate

NA= not applicable

Outcome measures

Work ability was measured via “The Work Ability Index” (WAI), a questionnaire of 7 items assessing the work ability of an individual employee by scoring from ‘2’ (very bad) to ‘10’ (very good). The WAI takes into consideration both the physical

and mental demands of the work. The index helps to identify those that need support. A point improvement on the WAI was denoted a substantial improvement and was used to demonstrate the effect for the cost effectiveness of the intervention.

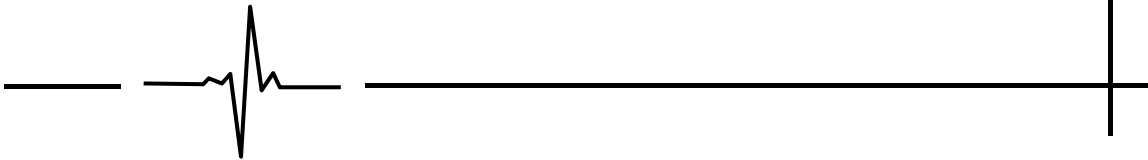
Differences in Quality Adjusted Life Years (QALYs) gained during the three month follow-up were assessed in the cost utility analysis. A QALY combines survival and utilities into one single measure. Utilities refer to the preference for any particular set of health outcomes and are indicated with a number between '0' (death) and '1' (perfect health). Utilities were calculated using EuroQol-5D-5L by measuring health outcomes on five dimensions (mobility, self-care, daily activities, pain/discomfort and depression/anxiety). Each dimension can be rated on five levels ranging from none to major complaints. A crosswalk to the 3 level value sets was used to calculate the utility weights based on EuroQol scores using Dutch tariff. Utilities at baseline and after three months follow-up were used to compute QALYs by means of the area under the curve method.

References technical appendix

- Oostenbrink JB, Koopmanschap MA, Rutten FF. Standardisation of costs: the Dutch Manual for Costing in economic evaluations. *Pharmacoeconomics*. 2002;20(7):443-454.
- Healthcare institute Netherlands. Pharmacotherapeutic compass [in Dutch: College van Zorgverzekeringen. Farmacotherapeutisch kompas]. [Cited September 2012]; Available from: <http://www.fk.cvz.nl/>.
- Drug costs in the Netherlands [cited 2014]. Available from: <http://www.medicijnkosten.nl/>.
- Hakkaart L, Tan SS, Bouwmans, CAM. Manual for cost research. Methods and standards costs for economic evaluations in healthcare [in Dutch: Handleiding voor kostenonderzoek. Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg]. Rotterdam, the Netherlands: Instituut voor Medical Technology Assessment, Erasmus Universiteit Rotterdam, 2010.
- Tuomi K IJ, Jahkola A, Katajarinne L, Tulkki A. Work Ability Index. 2nd revised edn. Helsinki: Finnish Institute of Occupational Health. 1998.
- Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011 Dec;20(10):1727-1736.
- EQ-5D User Guide. Rotterdam Erasmus Univeristeit Rotterdam, centrum voor gezondheidsbeleid en recht, 1995.
- Lamers LM, McDonnell J, Stalmeier PF, Krabbe PF, Busschbach JJ. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Econ*. 2006 Oct;15(10):1121-1132.
- EuroQol. [cited 2014]. Available from: <http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-value-sets.html>.

Part III

INNOVATIVE INTERVENTIONS



CHAPTER

8

Economic evaluation of a new organizational RTW intervention to improve Cooperation between Sick-listed employees and their Supervisors (COSS): A field study

Cindy Noben
Nicole Hoefsmid
Silvia Evers
Angelique de Rijk
Inge Houkes
Frans Nijhuis

First revision: Journal of Occupational and Environmental Medicine

Abstract

Objective: To assess the cost effectiveness, -utility, and -benefit of a new organizational return-to-work (RTW) intervention to improve Cooperation between Sick-listed employees and their Supervisors (COSS).

Methods: A field study with six months follow-up comparing COSS with common practice randomized participants aged 18-60, working at least 12 hours/week and absent for at least two weeks. Outcomes were initial RTW, quality adjusted life years (QALYs), and productivity gains.

Results: After six months, COSS generated less costs when compared with common practice. Participants in the COSS group returned to work earlier, improvement in QALYs were uncertain. Net benefits of COSS versus common practice yielded a productivity gain of €395.89.

Conclusion: Implementing COSS for sick-listed employees has potentials to reduce costs and improve productivity, and potentially quality of life. Longitudinal research might detect whether COSS also has the potential reaching sustainable RTW.

Key words: cost effectiveness analysis, cost utility analysis, cost benefit analysis, return-to-work, initial return-to-work, productivity, QALY, intervention, collaboration, employer, worker.

First revision: Journal of Occupational and Environmental Medicine

Introduction

Long-term sickness absence contributes largely to the total costs of work disability in Western countries. In the Netherlands, sickness absence among employees was up to 3.9 percent in 2013 [1]. Long term sickness absence adds to health- and economic problems as it is associated with a reduced probability of eventual return-to-work (RTW), followed by economic and social withdrawal. Reducing work disability to decrease absence and enhance RTW is an international government's top priority [2]. Cost effective RTW interventions are needed to achieve these aims. However, insight into the economic consequences of most existing RTW interventions is lacking. Although several studies performed cost effectiveness evaluations, most were among patients with specific diagnoses (i.e. distress, back pain, etc.) and are usually accessible to employees via healthcare providers or insurers [3-7]. These interventions can be regarded as a continuation of medical/clinical treatment (see, for example [8]).

The newly developed organizational intervention to enhance 'COoperation regarding return-to-work between Sick-listed employees and their Supervisors' (COSS) is innovative and generic as it is can be applied to all employees, regardless of their diagnoses or medical treatment. Supervisors will be interested in such interventions as they are not always allowed to ask employees about their health complaints [9]. COSS aims to resolve bottlenecks at the organizational level that were described in earlier research such as a lack of trust and open communication between the sick-listed employee and the supervisor [10-13]. COSS is complementary to Dutch legislation, which in itself appears to be insufficient to realise an adequate cooperation between sick-listed employees and their supervisors [13].

The objective of this study is to conduct an economic evaluation of COSS by assessing the cost effectiveness, -utility, and -benefit of COSS compared with common practice.

Methods

Study design

An economic evaluation was carried out alongside a field study comparing COSS with common practice with a six months follow-up. The field study was conducted in the Netherlands between 2012 and 2014 by researchers from Maastricht University. The field study whereof data was used for the economic evaluation was registered at the Dutch trial register (NTR3151). The Medical Ethics Committee (METC) of the Maastricht University was consulted and declared no ethical approval was necessary. The study protocol for conducting the economic evaluation and interpretations of the data [14] as described in this manuscript was approved by the local

Ethics Committee of Psychology at Maastricht University, the Netherlands (ECP107).

Most data were self-reported, except for sick-leave data which were retrieved from the participating organization's administration (a banking organization, which preferred to be anonymous).

An economic evaluation was conducted from both the societal and organizational perspective. Cost effectiveness- (CEA), cost utility- (CUA), and cost benefit (CBA) analyses were performed. The costs included were categorized into intervention, healthcare, patient and family, and other sectors (i.e. productivity). To avoid double counting in the CEA when productivity served the outcome of interest, solely intervention-, health care-, and patient and family costs were accounted for. More details on the cost categories can be seen in Table 2. (Supporting information 1 presents an overview of the different perspectives and types of economic evaluations conducted in this study).

Study population

Eligible participants were aged 18-60; sick listed for at least two weeks and performed paid labor for at least 12 hours per week at the onset of absence. Participants were excluded when their actual sick leave at baseline overreached 10 weeks or were terminally ill.

Interventions

Common practice

Dutch legislation requires supervisors to compensate their employees at sick-leave for at least 70% of their income during two years [9, 15, 16]. The Improved Gatekeeper Act prescribes obligatory procedures for employees and supervisors during this period (e.g. compose an action plan for RTW, monitor its progress, etc.) [17]. According to the Act, occupational physicians should be contacted (and paid for) by supervisors to analyze the employees' functional limitations and to advise regarding RTW [9, 15, 16].

Additional to legislation, the banking organization has an own sick leave policy that prescribed weekly meetings between sick-listed employees and their supervisors.

COSS

Additional to the legal requirements and the organizational policy, the cooperation between sick-listed employees and their supervisors was guided by the COSS intervention. COSS consisted of: (i) a conversation roadmap (step-by-step cooperation plan); (ii) a regular monitoring of the cooperation by using questionnaires at predetermined time points; and (iii) extra occupational physician support. The research

team analyzed the questionnaire results (cut-off scores) and wrote an advice on the cooperation for the OP who provided extra support if needed. The intervention was provided continuously until full RTW.

More in depth details regarding COSS- and common practice can be found elsewhere [18].

Main outcomes

The main outcomes were initial return-to-work (iRTW) and quality-adjusted life years (QALYs). iRTW was measured as duration in calendar days from the moment of employee inclusion onwards until initial return-to-work (iRTW) until maximally six months after inclusion. iRTW is conceptualized as the first moment at which progress is made in hours worked, even if the employee works less hours than contracted or if the employee is still (partly) sick listed [19]. Quality Adjusted Life Years (QALY) was assessed by using the self-administered EuroQol-5D-5L questionnaire. Health outcomes were measured on five dimensions: mobility, self-care, daily activities, pain and discomfort, and depression and anxiety [20]. Utilities were delineated by using the EQ-5D-5L crosswalk value set [21] to obtain Dutch tariffs and an overall utility score for population-based quality of life [22]. QALYs were calculated by multiplying the utility with the amount of time a patient spent in a particular health state. The maximum possible number of QALYs within the six months is 0.5 (half a year multiplied by 1, the optimal health state).

Resource use and valuation

Intervention costs

Intervention costs were calculated based on a bottom-up micro-costing approach (Supplementary file 2) and consisted of expenditures related to both developing and offering the intervention. For the economic evaluation, the costs of offering COSS were relevant; i.e. monitoring cooperation based on the questionnaires and training and extra support by occupational physicians (Supplementary file 2, part B). The reach (of potential participants) is based on the average number of sick-listed employees in the participating organization (N=2,266).

Healthcare costs

Self-reported questionnaires with a recall period of six months were posted to the participant at baseline and six months after randomization to collect resource use frequencies. Discounting of costs was not necessary because the follow-up was one year. At every assessment, information was collected about the volume of health care consumption using a modified version of the Trimbos/iMTA questionnaire for costs (Tic-P) [23]. The modified questionnaire comprised medical aids through consults and admissions to the general practitioner, occupational physician, move-

ment therapist, medical specialist, mental healthcare worker, alternative treatment, professional home care, hospitalization, rehabilitation clinic, and psychiatric institution. The respective units of resource use were multiplied by standard Dutch unit prices according to the Dutch guideline for costing research [24]. Costs were indexed to 2012 Euro (€) by means of the consumer price indexes of Statistics Netherlands [25]. Tariffs or average prices were used when standard price weights were not available. The costs of prescription medicines were calculated by deriving the standard daily dosage as published by the Dutch Pharmacotherapeutic Compass [26] and multiplied by the price per dose [27] and the number of intake days. The pharmacist's dispensing costs (€5.99) and the general practitioners' prescription costs were added (€14) to the costs of prescription medicines [24]. Over the counter drugs (including 6% taxes) were based on their market prices. Aids such as home interior elements, orthotics, mobility aids, etc. were based on the total costs per user for the reference year 2012 within the aid category as provided by the Dutch care institute [28].

Patient and family costs

Patient and family costs relate to informal care costs and parking and travel costs incurred while using health care resources. As the official unit price for informal care is not available, it was valued against the 'shadow price' of the wage rate per hour of a housekeeper [29]. Travelling and parking costs were calculated by multiplying average distance with the standard price weights according to the Dutch guideline for costing research or on average distance as presented in Warmerdam et al. 2010 [30].

Costs in other sectors (productivity costs)

The 'friction costs approach' (FCA) was used to value productivity costs. FCA states there is a short-run friction period during which the organization might incur losses as an adjustment is made to an employees' absence. However, in the long run, no (more) losses are held to occur because the employee either returned to work and the productivity returned, or the organization replaced the injured employee and productivity became comparable to what it was before [29, 31]. The friction cost approach thus assumes productivity losses only occur during the 'friction period'; the time needed to replace and train new employees. The friction period for the Netherlands was estimated based on the availability of labor—that is, the level of unemployment [32]. The friction period for 2012 in the Netherlands contained 13.24 weeks (92.68 days). Productivity cost estimates were based on self-reported average daily wage rates multiplied with the number of days until initial work resumption (sick-leave data retrieved from the banking organization) and with 0.8 to adjust for the elasticity between working time and productivity, which represents reduced labor efficiency, as prescribed by the costing manual [24]. Productivity costs were only calculated when lost work was not compensated (i.e. through extra working hours, by colleagues or by temporary or permanent replacement) as it was

assumed that absenteeism compensated for did not result in lost productivity. Detailed information on the calculation used to estimate the friction period and the productivity costs can be derived via the first author.

Data analyses

Data were analyzed according to the intention-to-treat principle. All participants were analyzed in the condition to which they were randomized after inclusion based on their baseline questionnaire responses. Missing values at baseline were imputed via mean imputation techniques. Missing data at follow-up were imputed by using the last observation carried forward, assuming no changes in effect or cost over time. All analyses were conducted in IBM SPSS Statistics 21 and Microsoft Excel.

Economic Evaluation

For the cost effectiveness analyses (CEA), incremental cost effectiveness ratios (ICERs) were determined to indicate the additional investments needed for COSS to gain one extra unit of effect compared with common practice. The following calculation was used; $(C_1 - C_0) / (E_1 - E_0)$, where C is the average cost and E is the proportion of respondents achieving initial RTW. The total costs were calculated by the sum of the intervention-, healthcare-, and patient and family costs. In order to calculate the incremental cost utility ratio (ICUR) the incremental costs (intervention-, healthcare-, patient and family-, and productivity costs) are divided by the difference in QALYs. The cost benefit analysis (CBA) calculated the net monetary benefit (NB) and the benefit-cost ratio (BCR) from the organizational perspective. The NB was derived by subtracting the mean difference in intervention costs between both groups from the difference in productivity costs. The BCR was calculated by dividing the benefits of the investment by the costs of the investment in order to calculate the amount of money returned per Euro invested. Financial returns were positive if $NB > 0$ and $BCR > 1$.

Non-parametric bootstrap re-sampling techniques were used to capture the uncertainty surrounding the incremental costs. Using the original data, 5,000 replications of the cost-effect pairs were generated and plotted on one of the four quadrants of the cost effectiveness plane to display differences in costs on the vertical axis and differences in effects on the horizontal axis. In the North-east quadrant the intervention produces superior effects at additional costs relative to common practice. In the North-west quadrant less effect is produced for additional costs meaning the intervention is “dominated” by the common practice condition. In the South-west quadrant less effect is produced, but there are some cost-savings. And finally, in the South-east quadrant the intervention generates superior effects for less costs; the intervention “dominates” common practice.

The bootstrapped ratios were furthermore used to construct an acceptability curve which indicates the probability of the intervention being cost effective at a

predetermined ceiling ratio (€); which is society's willingness to pay (WTP) to gain one extra unit of effect (e.g. one QALY). The WTP depends on the type of disease and its severity; which is higher for very severe conditions (i.e. a tentative maximum of €80,000) than for mild conditions in the Netherlands [33]. For the COSS intervention it was difficult to choose one threshold based on the severity of conditions, as participants in the field study have a variety of diseases. The acceptability curve however demonstrates the probability of COSS being cost effective at a ceiling ratio of €18,000, indicating that COSS has the likelihood of being cost effective for the whole spectrum of diseases from non-severe till severe conditions (see Figure 2, part B).

To ascertain the robustness of the findings, different sensitivity analyses were carried out. As elder employees might stay on longer sick leave or use the claim of work disability absenteeism as an alternative to retirement, especially among those over 55 years of age was used [34], the first sensitivity analysis therefore tested whether faster work resumption due to COSS can be achieved among employees between 40 and 60 years of age when compared to common practice. Secondly, because COSS is developed to be broadly implementable because of its non-disease specific focus, a subgroup analysis controlling for chronic disability is conducted. Thirdly, a cost utility analysis was calculated by using similar costs as the cost effectiveness analysis. Finally, a cost benefit sensitivity analysis was performed in which the intervention costs were higher as they were based on the actual uptakes (n=39).

Results

Participants

90 participants from six departments at the banking organization were randomized to the COSS- and common practice group and filled in the baseline questionnaire. Complying with the inclusion criteria, the data of 39 participants in the COSS group and 25 in the common practice group were analyzed at baseline and after six months follow-up. Figure 1 provides more details on the participants' flow throughout the study.

Figure 1. Participant's flow chart

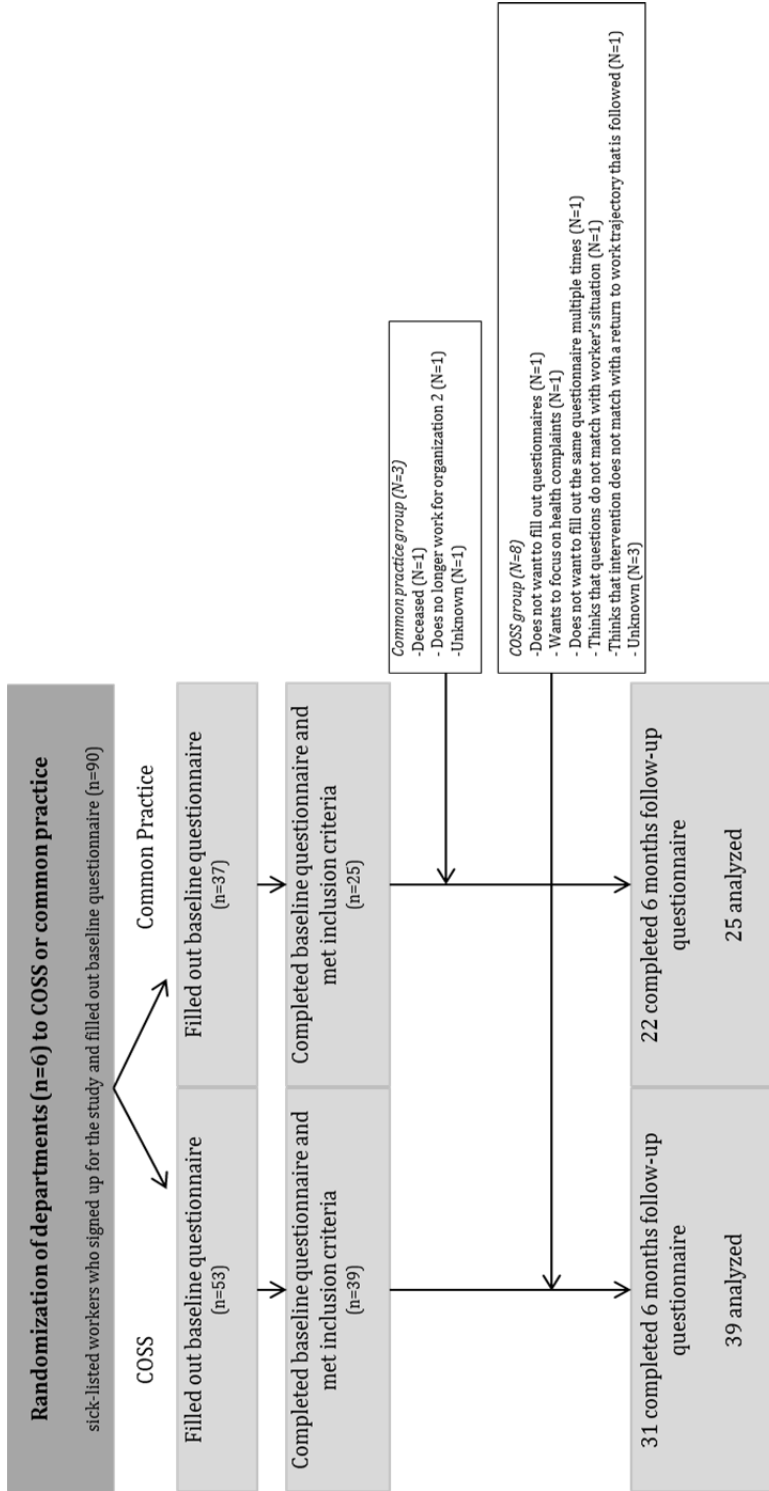


Table 1 lists the baseline characteristics of both groups. Significant differences across both groups were found for age, gender and the average number of working hours. None of the cost categories were significantly different when comparing both groups, though on average, costs were higher in the common practice group.

Table 1. Baseline characteristics (n=64)

	Intervention (n=39)	Group	Control (n=25)	Group	p-value
Age, mean (sd)	45.31 (9.2)		50.6 (7.4)		0.019 (a) *
Female, N (%)	25 (64)		7 (28)		0.005 (b) **
Education, N (%)†					
Low	7 (18)		6 (24)		
Intermediate	19 (49)		4 (16)		
High	12 (31)		14 (56)		0.06 (b)
Other	1 (2)		1 (4)		
Working hours, mean (sd)	31.72 (6.8)		35.50 (4.6)		0.028 (a) *
Marital Status, N (%)					
Married	24 (62)		21 (84)		
Living together with a partner	9 (23)		3 (12)		
Divorced	5 (13)		0		0.15 (b)
Single	1 (2)		1 (4)		
Children under 12 years, N (%)	18 (46)		6 (24)		0.07 (b)
Permanent employment, N (%)	39 (100)		24 (96)		0.21 (b)
Managerial position, N (%)	3 (8)		4 (16)		0.3 (b)
Shift work, N (%)	2 (5)		1 (4)		0.84 (b)
General Health, N (%)					
Excellent	1 (3)		1 (4)		
Very Good	10 (26)		7 (28)		
Good	22 (56)		14 (56)		0.97 (b)
Moderate	6 (15)		3 (12)		
Long term disability, N (%)					
None	30 (77)		16 (64)		
One or more	9 (23)		9 (36)		0.26 (b)
EQ-5D-5L Utilities, mean (sd)‡	0.64 (0.16)		0.63 (0.24)		0.8 (a)
Costs, mean in € (sd)					
Healthcare costs	1,348.56 (1,609.7)		1,813.64 (2,099.89)		0.72 (c)
Medication & aids	110.1 (206.17)		125.6 (184.48)		0.24 (c)
Health care service use	1,238.46 (1,454.03)		1,688.04 (2,017.45)		0.7 (c)
Patient and family costs	208.03 (376.06)		254 (462.18)		0.91 (c)
Informal care costs	137.12 (364.54)		190.03 (457.03)		0.85 (c)
Travel and parking costs	70.9 (60.48)		63.77 (40.32)		0.92 (c)

†Low=preschool or primary school; intermediate=lower and upper secondary; high=tertiary school, university, or postgraduate | ‡High scores indicate better perceived health outcomes | (a) Anova (b) Pearson Chi-squared test (c) Non-parametric Kruskal Wallis test | * Significant at the 5% level ** Significant at the 1% level

Costs

After six months follow-up, no significant cost differences were found (Table 2). Despite for intervention costs, the average costs per category were highest in the common practice group. Occupational physician-, mental healthcare worker-, alternative treatment-, hospitalization-, and prescribed medication costs were higher in the COSS group. Incremental costs based on the bootstrapped subtotals indicated costs were lower in the COSS group when compared to the common practice group.

Table 2. Total costs per cost item and subtotal bootstrapped costs after six months follow-up

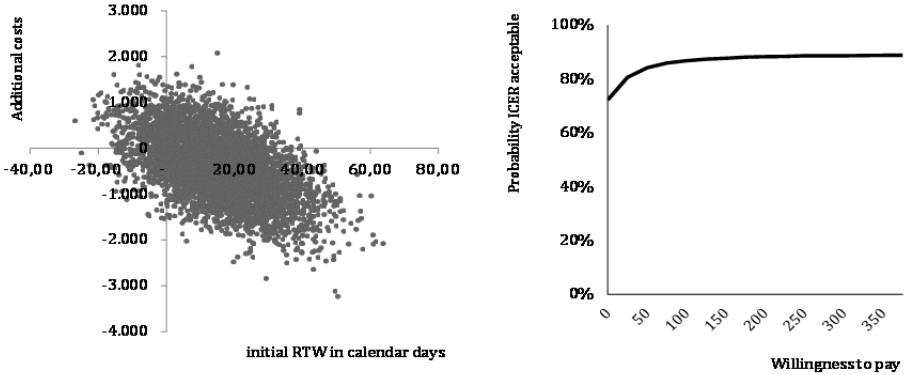
Cost category	Intervention group, costs € (n=39)	Common practice group, costs € (n=25)	Incremental costs, € (2,5-97,5 percentile)
Intervention costs	2.19	0	n.a.
Healthcare costs			
General practitioner	90.71	98.7	
Occupational physician	140.9	112.88	
Movement therapist	236.35	367.39	
Medical specialist	100.35	187.32	
Mental healthcare worker	243.05	58.9	
Alternative treatment	45.89	13.42	
Professional home care	8.85	8.28	
Rehabilitation clinic	0	0	
Psychiatric institution	0	0	
Hospitalization (day parts)	314.38	191.91	
Hospitalization(admissions)	783.76	1308.47	
Prescription medication	77.21	50.6	
Over-the-counter medication	4.79	10.64	
Aids	3.22	40.88	
Subtotal	2049.47	2449.39	
Bootstrapped subtotal, mean (sd)	1105.57 (140.76)	1710.51 (330.81)	-605 (-1343 to 38)
Patient and family costs			
Informal care	106.5	114.12	
Parking and travel costs	94.34	110.75	
Subtotal	200.84	224.88	
Bootstrapped subtotal, mean (sd)	201.49 (50)	226.34 (59.24)	-25 (-175 to 129)
Costs in other sectors			
Productivity costs	1004.64	1918.89	
Bootstrapped subtotal, mean (sd)	1078.03 (288.77)	1944.80 (497.05)	-867 (-1991 to 187)

Cost effectiveness analyses

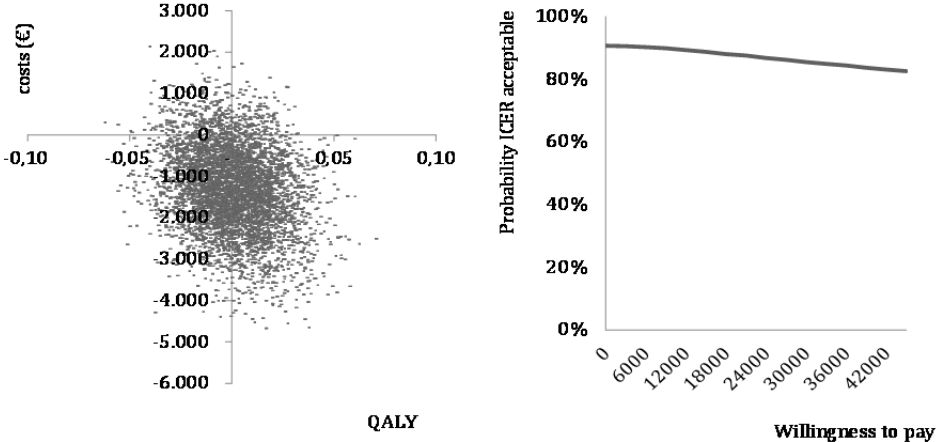
The CEA (see Table 3) showed an ICER of 26.78 ($\Delta C = -421.77 / \Delta E = 15.75$) meaning COSS was less costly and more effective (i.e. almost 16 reduced calendar days of absenteeism) compared to common practice and €26.78 is saved per day of work resumption. See also Figure 2 (part A) where 68% of the bootstrapped ICERs are located in the south east quadrant (indicating more effect and less costs). The cost effectiveness acceptability curve (CEAC) indicates that at a willingness to pay of €100 for one day earlier work resumption, the probability of COSS being cost effective is 87%.

Figure 2. Cost effectiveness plane and cost effectiveness acceptability curve for differences in (A) initial RTW, (B) QALYs

(A)



(B)



Cost utility analysis

The bootstrapped results in Figure 2 (part B) show that 44% and 47% of the cost effectiveness pairs are located in respectively the southwest and southeast quadrant where COSS is less costly but the difference in effect is uncertain (wide spread of the point estimates, Table 3). Based on the acceptability curve, COSS has a probability of approximately 88% of being cost effective at the considered ceiling ratio of €18,000 for this study. The acceptability curve descends, meaning that the probability of COSS being cost effective decreases when the value society attaches to a QALY increases.

Table 3. Mean cost and effect differences between the COSS intervention group and common practice

Main analyses	Outcome	COSS					Common Practice				
		<i>Sample size</i>	<i>Costs, €</i>	<i>Effect</i>	<i>Sample size</i>	<i>Costs, €</i>	<i>Effect</i>	Δ <i>Cost</i>	Δ <i>Effect</i>	<i>ICER/ICUR (95% CI)</i>	
Cost Effectiveness	Initial RTW (calendar days)	39	2,252.50	43.77	25	2,674.27	59.52	-421.77	15.75	-26.78 (-300 to 256)	
Cost Utility	QALY	39	3,257.14	0.3587	25	4,593.16	0.3584	-1,336.02	0.0003	-4,202,013.55 (-1,293,021 to 1,260,056)	
Sensitivity analyses – Cost effectiveness analyses											
Aged between 40 and 60	Initial RTW (calendar days)	27	2,129.44	41.74	22	2,809.56	58.36	-680.12	16.62	-40.91 (-318 to 281)	
Chronic disease	Initial RTW (calendar days)	9	2,772.63	41	9	2,414.52	60.11	358.11	19.11	18.74 (-543 to 440)	
No chronic disease	Initial RTW (calendar days)	30	2,096.46	44.60	16	2,820.38	59.19	-723.92	14.59	-49.63 (-457 to 384)	
Sensitivity analyses – Cost utility analysis											
Omitting productivity costs	QALY	39	2,252.50	0.3587	25	2,674.27	0.3584	-421.77	0.0003	-1,326,528.13 (-546,492 to 632,170)	

Cost benefit analysis

The costs of the intervention were €2.19 higher than in common practice. Productivity costs according to the friction cost method were €867 lower in the COSS group compared to the common practice group (Table 2). The net benefit for the supervisor is €864.81 and the BCR is €395.89; indicating a productivity gain for the supervisor of almost €396 per invested Euro.

Sensitivity analysis

With respect to the first sensitivity analysis, results were robust when using the data of employees aged between 40 and 60 years. Almost €41 was saved per work resumption day. Results of the main analysis were also similar to the sensitivity analysis whereby only employees without a chronic disability were included in the analysis. One day earlier work resumption among employees without a chronic disability resulted in a cost saving of almost €50. The subgroup of employees with a chronic disability yielded faster return to work, however, contrasting the main analysis; an additional investment of almost €18 was needed. A summary of the findings is shown in Table 3. With respect to costs per QALY, the results of the fourth sensitivity analysis were robust when omitting productivity costs. The acceptability curve descended from 74% (best case with 3,000 Euro willingness to pay to gain one QALY) to 65% (worst case). Assessing the robustness of the findings from the organizational perspective by accounting the costs of the intervention depending on the actual uptake rates (N=39) were comparable to the results of the main analysis. The intervention costs increased from 2.19 Euro to 127.35 Euro. The net benefit for the supervisor is €739.65 and the BCR productivity was €6.89 per invested Euro, indicating that COSS still produced a positive financial return.

Conclusions

A new organizational RTW intervention for Cooperation between Sick-listed employees and their Supervisors (COSS) appeared to have substantial economic benefits over common practice. The cost effectiveness acceptability curve indicated that the probability of COSS being cost effective compared with common practice was 87% for ceiling ratios larger than €100 to gain one day of earlier initial return to work. For QALYs gained, the cost effectiveness acceptability curve indicated that COSS had a cost effectiveness probability of almost 88% at a ceiling ratio of €18,000.

This study combined different types of economic evaluations (i.e. cost effectiveness, -utility, and -benefit) from diverse perspectives to compare the organizational intervention, regardless of the employees' diagnoses, with common practice. Lower costs and higher effects of COSS when compared to common practice result-

ed in an ICER of almost €27 per initial return-to-work calendar day. Exploring the robustness of the results in several sensitivity analyses resulted in comparable findings. COSS remained cost effective, even among the group of older employees (age range between 40 and 60 years) that generally have longer sickness absence spells [34]. Employees either with or without a chronic illness receiving COSS returned to work faster when compared to common practice. Results in terms of QALYs were uncertain because the majority of the incremental cost utility ratios were located in the quadrant of ‘fewer costs, more effects’ and ‘fewer costs, less effects’. Exploring the incremental cost utility omitting productivity costs resulted in comparable findings. From an organizational perspective, COSS can be regarded as cost saving to the employer, even when upgrading the intervention costs. A recent review yielded similar overall positive results on the financial return of worksite health promotion programs [35].

Strengths and limitations

The main strengths of this study were its pragmatic (‘real life setting’) design, the different perspectives (i.e. societal, organizational, healthcare) used to assess different types of economic evaluations (CEA, CUA, and CBA), the use of objective sick-leave data, and the transparent registration of the intervention costs, limiting the risk of over- or underestimation actual costs. Another strength is the inclusion of productivity loss caused by reduced labor efficiency by accounting for elasticity when using the net cumulative days of sick leave as the proxy for productivity loss. Accounting for presenteeism reduced the risk of over- or underestimating actual productivity costs in both groups alike. The effectiveness study confirmed the findings of this study when investigating the time to first RTW (i.e. a tendency towards effectiveness in the COSS group). Also, when addressing full- and sustainable RTW in calendar days, positive effects were found in favor of the COSS group [36]. Further research is needed to detect whether COSS has the potential of reaching sustainable RTW at lower costs when compared to common practice. To economically evaluate the potential positive effects of COSS on full- and sustainable RTW over an adequate follow-up period and a longer time span more research is needed as it took more time than anticipated to recruit individual employees and supervisors. Furthermore, the generalizability of the study results is not boundless. The Netherlands holds a unique position regarding RTW policy [16], and COSS was specifically tailored to and implemented in the Dutch context. Taking into consideration that the common practice is already more intensive due to legal requirements for cooperation [16] when compared to other countries (e.g. Belgium, Canada or the US), the effects of the intervention might be higher in countries where no legal requirements for cooperation and collaboration are set.

Concluding remarks

In the Netherlands, there is a strong need for research regarding cost effectiveness studies of interventions on working conditions, social work environment and health [37]. This study demonstrated the cost effectiveness of COSS. For the cost utility, the high utility scores at baseline might explain the uncertainty regarding the effectiveness for both COSS and common practice, making it difficult to detect an increase over time. COSS yielded a cost saving and from an organizational perspective even a positive financial return regarding first RTW.

Acknowledgments

We want to thank the banking organization and the study participants for their collaboration and participation. We also thank medical doctor Kees de Kock for his assistance while estimating the per person average number of days of medication usage based on time spans of usage, type and combinations.

References

1. Statistics Netherlands. Ziekteverzuim op laagste punt sinds 1996 [Sickness absence at lowest level since 1996]. 2014 [cited 2014]; Available from: <http://www.cbs.nl/en-GB/menu/themas/arbeid-sociale-zekerheid/publicaties/artikelen/archief/2014/2014-4047-wm.htm>.
2. Organisation for Economic Co-operation and Development. Sickness, Disability and Work: breaking the barriers. 2010 [cited 2014]; Available from: http://ec.europa.eu/health/mental_health/eu_compass/reports_studies/disability_synthesis_2010_en.pdf.
3. van Oostrom SH, Heymans MW, de Vet HC, van Tulder MW, van Mechelen W, Anema JR. Economic evaluation of a workplace intervention for sick-listed employees with distress. *Occup Environ Med*. 2010 Sep;67(9):603-10.
4. Loisel P, Lemaire J, Poitras S, Durand MJ, Champagne F, Stock S, et al. Cost-benefit and cost-effectiveness analysis of a disability prevention model for back pain management: a six year follow up study. *Occup Environ Med*. 2002 Dec;59(12):807-15.
5. Vermeulen SJ, Heymans MW, Anema JR, Schellart AJ, van Mechelen W, van der Beek AJ. Economic evaluation of a participatory return-to-work intervention for temporary agency and unemployed workers sick-listed due to musculoskeletal disorders. *Scand J Work Environ Health*. 2013 Jan;39(1):46-56.
6. Rebergen DS, Bruinvels DJ, van Tulder MW, van der Beek AJ, van Mechelen W. Cost-effectiveness of guideline-based care for workers with mental health problems. *J Occup Environ Med*. 2009 Mar;51(3):313-22.
7. Noben C, Smit F, Nieuwenhuijsen K, Ketelaar S, Gartner F, Boon B, et al. Comparative cost-effectiveness of two interventions to promote work functioning by targeting mental health complaints among nurses: pragmatic cluster randomised trial. *Int J Nurs Stud*. 2014 Oct;51(10):1321-31.
8. Tamminga SJ, Verbeek JH, Bos MM, Fons G, Kitzen JJ, Plaisier PW, et al. Effectiveness of a hospital-based work support intervention for female cancer patients - a multi-centre randomised controlled trial. *PLoS One*. 2013;8(5):e63271.
9. Overheid.nl. Wet Verbetering Poortwachter (Improved Gatekeeper Act). Available from: http://wetten.overheid.nl/BWBR0013063/geldigheidsdatum_10-04-2014.
10. Friesen MN, Yassi A, Cooper J. Return-to-work: The importance of human interactions and organizational structures. *Work*. 2001;17(1):11-22.
11. Wrapson W, Mewse AJ. Supervisors' responses to sickness certification for an episode of low back pain: employees' personal experiences. *Disabil Rehabil*. 2011;33(19-20):1728-36.
12. Kremer AM, Steenbeek R. Avoidable sickness absence in a Dutch working population. *J Occup Rehabil*. 2010 Mar;20(1):81-9.
13. Hoefsmit N, de Rijk A, Houkes I. Work resumption at the price of distrust: a qualitative study on return to work legislation in the Netherlands. *BMC Public Health*. 2013;13:153.
14. Noben CY, Nijhuis FJ, de Rijk AE, Evers SM. Design of a trial-based economic evaluation on the cost-effectiveness of employability interventions among work disabled employees or employees at risk of work disability: The CASE-study. *BMC Public Health*. 2012 Jan 18;12(1):43.
15. Hop AD, de Pijper B, Sens IM, Wildenburg A, Baars MJH, Dekker EM, et al. De kleine gids voor de Nederlandse sociale zekerheid 2010.twee 81 ed. Hop AD, editor. Hoofddorp: Kluwer; 2010 1 juli 2010. 206p.
16. Elders L, Burdorf A. Occupational medicine in The Netherlands. *Occup Med (Lond)*. 2010 Jun;60(4):314.
17. Evers HGJ. De Wet Verbetering Poortwachter. Handleiding voor de praktijk (The Improved Gatekeeper Act. A practice guideline). Deventer: Kluwer Juridisch; 2007.
18. Hoefsmit N, Boumans N, Houkes I, Nijhuis F. A process evaluation of a return-to-work intervention to improve Cooperation between Sick-listed employees and their Supervisors (COSS). 2014.
19. De Rijk A, Janssen N, Alexanderson K, Nijhuis F. Gender differences in return to work patterns among sickness absentees and their associations with health: a prospective cohort study in The Netherlands. *Int J Rehabil Res*. 2008 Dec;31(4):327-36.

20. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res.* 2011 Apr 9.
21. EuroQol. EQ-5D-5L Value Sets [cited 2014]. Available from: <http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-value-sets.html>.
22. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff. [In Dutch: Kwaliteit van leven meten in economische evaluaties: het Nederlands EQ-5D-tarief]. *Ned Tijdschr Geneesk.* 2005 Jul 9;149(28):1574-8.
23. Hakkaart-van Roijen L, van Straten A, Donker M, Tiemens B. Manual Trimbos/iMTA questionnaire for Costs associated with Psychiatric illness(TiC-P) [In Dutch: Handleiding Trimbos/iMTA questionnaire for Costs associated with Psychiatric illness (TiC-P)]: Rotterdam: Instituut voor Medische Technology Assessmentm, 2002.
24. Hakkaart L, Tan SS, Bouwmans, CAM. Manual for cost research. Methods and standards costs for economic evaluations in healthcare [in Dutch: Handleiding voor kostenonderzoek. Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg]. Rotterdam, the Netherlands: Instituut voor Medical Technology Assessment, Erasmus Universiteit Rotterdam, 2010.
25. Statistics Netherlands. Consumer prices. 2014 [cited 2014]; Available from: [http://statline.cbs.nl/StatWeb/publication/?DM=SLen&PA=80087ENG&D1=0-1,4-5&D2=0&D3=\(1-39\)-l&LA=EN&VW=T](http://statline.cbs.nl/StatWeb/publication/?DM=SLen&PA=80087ENG&D1=0-1,4-5&D2=0&D3=(1-39)-l&LA=EN&VW=T).
26. Healthcare institute Netherlands. Pharmacotherapeutic compass [in Dutch: College van Zorgverzekeringen. Farmacotheurapeutisch kompas]. [Cited September 2012]; Available from: <http://www.fk.cvz.nl/>.
27. Healthcare Institute Netherlands. Medicijnkosten [Drug costs in the Netherlands]. [cited 2014]; Available from: <http://www.medicijnkosten.nl/>.
28. Healthcare institute Netherlands. GIP databank [GIP database]. [cited 2014]; Available from: www.gipdatabank.nl.
29. Oostenbrink JB, Koopmanschap MA, Rutten FF. Standardisation of costs: the Dutch Manual for Costing in economic evaluations. *Pharmacoeconomics* 2002;20:443-54.
30. Warmerdam L, Smit F, van Straten A, Riper H, Cuijpers P. Cost-utility and cost-effectiveness of internet-based treatment for adults with depressive symptoms: randomized trial. *Journal of medical Internet research.* 2010;12(5): e53.
31. Koopmanschap MA, van Ineveld BM. Towards a new approach for estimating indirect costs of disease. *Soc Sci Med.* 1992 May;34(9):1005-10.
32. Statistics Netherlands. Jobs (outstanding, new and filled). [cited 2014]; Available from: <http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=80857NED&LA=NL>
33. Council for Public Health and Health Care. Sensible and sustainable care (in Dutch). Zoetermeer: Council for Public Health and Health Care; 2006.
34. Hesselings JH, Houtman I, van den Bossche S. Trends in ziekteverzuim (Trends in sickness absence). The Hague, The Netherlands: Sociaal Cultureel Planbureau; 2012.
35. van Dongen JM, Proper KI, van Wier MF, van der Beek AJ, Bongers PM, van Mechelen W, et al. Systematic review on the financial return of worksite health promotion programmes aimed at improving nutrition and/or increasing physical activity. *Obesity reviews : an official journal of the International Association for the Study of Obesity.* 2011 Dec;12(12):1031-49.
36. Hoefsmit N, Houkes I, Boumans N, Noben C, Winkens B, Nijhuis F. The effectiveness of an intervention to enhance cooperation between sick-listed employees and their supervisors (COSS). Submitted: *Journal of Occupational Rehabilitation.*
37. Burdorf A. Research in occupational medicine in The Netherlands: responsiveness to societal needs. *Occup Med (Lond).* 2010 Jun;60(4):242-4.

Supporting information 1. Main analyses for economic evaluation

Method	Costs (€)	Outcome
Cost Effectiveness	- healthcare costs - patient and family costs - intervention costs	Initial RTW
Cost Utility	- healthcare costs - patient and family costs - intervention costs - costs in other sectors (productivity)	Quality Adjusted Life Years
Cost Benefit	- intervention costs	Monetary benefits (productivity gain)

Supporting information 2. Total per participant intervention costs

Type of costs	Calculation	Outcome (€)
Part A: Development costs		
Development of the conversation roadmap	Average wage rates per contributor multiplied with average number of hours worked on the developments	2,112.54
Development of questionnaires and online versions. Development of cut-off scores.	Average wage rates per contributor multiplied with average number of hours worked on the developments	2,447.64
Sending out questionnaires and reminders, reminder phone calls and questionnaire analyses for occupational physicians and coaches	Average wage rates per contributor multiplied with average number of hours worked	5,932.50
Improving COSS based on feedback	Average wage rates per contributor multiplied with average number of hours worked	158.20
Total costs:		5,311.88
Part B: Costs of offering the intervention		
Printing costs of the conversation roadmap	Number of prints multiplied with the amount per print	561.60
Use of roadmap by supervisors	Hours required for reading and using the complete roadmap multiplied with average sick leave duration in months multiplied with average wage p. hour supervisor	252.11
Monitoring cooperation: Filling out questionnaires by supervisors	Average time per questionnaire multiplied with average wage supervisor	814.96
Preparing and providing an accredited training session to occupational physicians/coach	Average wage rates per contributor multiplied with average number of hours worked	1,603.90
Occupational physicians and coaches providing extra support to employees and/or supervisors	Average wage rates per contributor multiplied with average number of hours worked	1,734.42
Total costs:		4,966.99
Costs per participant*		2.19
* 2,266 participants: representation of employees in the banking organization, accounting for the percentage of absence during the reference year 2012		

CHAPTER

9

Economic evaluation of an intervention program with the aim to improve work productivity for workers with rheumatoid arthritis: A six months interim analysis*

Cindy Noben
Myrthe van Vilsteren
Cécile Boot
Romy Steenbeek
Dirkjan van Schaardenburg
Han Anema
Silvia Evers
Frans Nijhuis
Angelique de Rijk

** the included economic evaluation results are based on interim-analyses as data collection is ongoing*

Abstract

Objective To evaluate the cost effectiveness and cost utility of an integrated care- and participatory workplace intervention provided to workers with rheumatoid arthritis (RA) to improve their work productivity.

Methods An economic evaluation was conducted alongside a randomized controlled trial with a six months follow-up in specialized rheumatology treatment centres in or near Amsterdam, the Netherlands. Adults aged 18-64 years, diagnosed with RA, working in a paid job for at least 8 hours per week and experiencing at least minor difficulties in functioning at work were randomized to the intervention (n=75) or the control group (n=75). Effect outcomes were at-work productivity loss (measured in presenteeism hours) and quality adjusted life years (QALYs). Health care costs, patient and family costs, costs in other sectors, and intervention costs were calculated from a societal perspective. Cost effectiveness- and cost utility analyses were conducted to indicate the incremental costs and benefits per additional unit of effect. A subgroup analysis and several sensitivity analyses to evaluate the impact of different methodological considerations on the results were conducted to attest the robustness of the findings.

Results At-work productivity loss was about 4.5 hours in the intervention group and 4 hours in the usual care group per two weeks (non-significant difference for both group, p-value=0.103). The mean QALY was on average 0.39 for the control group and 0.38 for the intervention group, which was not significantly different comparing both groups (p-value=0.09). In total, average costs after six months follow-up were highest in the intervention group (€1,969.48) compared to the control group (€1,456.13). The main cost effectiveness and cost utility analyses show that the intervention was more expensive and less effective than the control condition. Sensitivity analyses supported these findings.

Conclusions Based on the six months data no evidence is forehanded to support the implementation of the intervention on cost-related grounds.

Introduction

Rheumatoid arthritis (RA) patients experience more restrictions in participation in employment, have higher risk of becoming work disabled, and become more work disabled over time [1-3]. RA is a prevalent condition (between 0.5% and 1% of the population in the developed world) associated with severe impairments [4]. Furthermore, RA disability related productivity losses due to time lost from work or during work (respectively absenteeism and presenteeism) have a substantial socio-economic impact [5-7]. Previously conducted research on work disability in RA showed that over a period of almost 13 years, almost 9% of RA patients gave up their paid work and 4% stopped and never resumed paid work [8]. According to Boonen & Severens (2011), sickness absence among patients with early RA ranges between 53-82% per year, resulting in substantial mean annual cost per patient for paid productivity loss (up to about €8,000) [4]. Absenteeism from paid work represents a major source of productivity costs; the weighted mean annual sick leave costs were as high as €2,770 per patient [4]. Presenteeism, which refers to at-work productivity loss, is an important cost contributor as well. Furthermore, presenteeism can occur without absenteeism, it can precede absenteeism, and it can follow absenteeism [4, 9]. Particularly in situations where the employee is not absent from paid work but struggling with work- and health related disability complaints, as is often the case with RA patients, costs related to presenteeism are at least as important as absenteeism costs.

In an attempt to reduce the burden for patients with RA regarding their work abilities, and to reduce high costs to society, a need emerged to develop (cost) effective worksite interventions to support work participation of workers with RA. Previously conducted studies have shown that suboptimal care by lack of communication may lead to long-term absenteeism from work and permanent work disability, and thereby high costs [10]. Integrated care programs on work participation in chronically ill patients were therefore developed, inducing that integrated care might increase work participation [11, 12]. Because of the need for recommendations for work activities in chronic diseases and the fact that involving the worksite might make a difference, an intervention program consisting out of both integrated care and a participatory workplace intervention was provided to workers with RA. Integrated care was already effective in increasing work participation among chronically ill patients [11, 12]. An integrated care program for patients sick listed because of chronic low back pain also had substantial economic benefits over usual care [11]. A participatory return-to-work intervention for sick-listed workers due to musculoskeletal disorders enhanced work resumption and generated a net socio-economic benefit [13].

Before implementing this intervention on a large scale, a broader picture of the total cost and effects of the intervention needs to be evaluated. As (occupational) healthcare budgets are limited, economic evaluations become more important. Cost

effectiveness analyses are conducted to estimate the incremental monetary costs and benefits of an intervention per unit of effect gained. Although intervention related costs to evaluate the cost effectiveness of a worksite intervention are important to account for, worksite interventions are also associated with healthcare and productivity-related costs. There is still an ongoing debate on how to identify, measure, and value productivity costs in economic evaluations [9, 14].

Economic evaluations are recommended to be conducted from the perspective related to the costs being evaluated [15]. According to the guidelines for economic evaluations, full economic evaluations are preferred to be conducted from the societal perspective, considering all costs and benefits of interventions no matter on whom they fall [16]. Adhering to the guidelines, this study includes all relevant costs from the societal perspective (healthcare-, patient and family-, and costs in other sectors). Defining costs in other sectors, this study focused on costs related to productivity. The main productivity cost terms and the different conceptions on how to value productivity costs as presented throughout this study are listed in textbox 1 (adapted from Krol, et al. 2013) [14].

Textbox 1. Main productivity cost terms

Absenteeism	Not attending work
Presenteeism	Attending work with diminished functioning (reduced work quantity and quality)
Compensation	Methods in which lost productivity is compensated for (e.g. extra hours, co-workers took over, etc.)
Human Capital approach	All potential lost productivity not performed by a person due to work disability
Friction Cost approach	Lost productivity is limited until the time needed to replace a work disabled worker and train his/her substitute

The aim of this present study was to conduct an economic evaluation, including cost effectiveness- and cost utility analyses, in which the intervention (i.e. an intervention program consisting out of both integrated care and a participatory workplace intervention) was compared to usual care of workers with RA. The research questions were:

- (i) "Is the joint distribution of costs and at-work productivity loss preferable in the intervention group when compared to usual care?"
- (ii) "Is the joint distribution of costs and RA patients' quality of life preferable in the intervention group when compared to usual care?"
- (iii) "How do different methodological considerations under a given set of assumptions impact the results? (risk-case analysis and sensitivity analyses)"

Methods

The economic evaluation was conducted alongside a randomized controlled trial comparing a workplace intervention, which is consisting out of both integrated care and a participatory workplace intervention, with care as usual among workers with rheumatoid arthritis (RA). The study was conducted in the Netherlands between 2011 and 2013 and the follow-up period was six months. The trial was registered in the Dutch Trial Register (NTR2886). The Medical Ethics Committee of the Slotervaart Hospital and Reade, and the VU University Medical Center, Amsterdam, the Netherlands approved the study design, protocols, procedures, and informed consent. Participation was voluntary and all participants signed informed consent.

Study population

RA patients who have visited a rheumatologist of one of the participating hospitals during the last year were approached. Eligible patients were adults aged 18-64 who were diagnosed with RA, had a paid job (paid-employment or self-employed) for at least 8 hours per week and experienced at least minor difficulties in functioning at work. Patients were excluded in case of (i) severe comorbidity as it would hamper compliance to the protocol; (ii) being unable to read or understand Dutch language, and (iii) taking more than three months sick leave at time of inclusion. Eligible patients received an information letter about the project from their own rheumatologist. Details of the study design are described elsewhere [17].

Sample size

At-work productivity loss, expressed as hours lost from work due to presenteeism, measured with the Work Limitation Questionnaire (WLQ) [18], was used to calculate the sample size. A two hour per two weeks was considered a relevant difference, which implies a moderate standardized effect of 0.5. Assuming a dropout rate of 15%, power analysis revealed a sample size of 142 patients to be included in total (80% and an alpha of 0.05) [17].

Randomization

Randomization to either the intervention group or control group was performed on patient level. To prevent unequal randomization, patients were pre-stratified by three prognostic factors: sex, number of working hours per week, and whether the patient performed light or heavy physically/ mentally demanding work [19]. A software program (Minim) based on the minimisation method was used for treatment allocation [20]. Minimisation allows pre-stratification by several prognostic factors, even in small samples [21, 22]. Due to the character of the intervention,

patients, therapists and researchers could not be blinded for the allocated treatment after randomization.

Intervention

The participants in the intervention group were able to take part in the intervention program consisting out of two components which complement each other: (i) integrated care, and (ii) participatory workplace intervention. Integrated care was provided based on a case management protocol whereby a trained clinical occupational physician (COP) acted as care manager. The care manager had an intermediate role between clinical and occupational care. During the first consultation the COP was responsible for history taking and physical examination, the planning and coordination of care, and for communication between all care providers. The patient visited the care manager within one week after randomization and, if needed, again after six and 12 weeks to evaluate and if necessary adjust the treatment plan. The participatory workplace intervention concerned workplace adaptations based on active participation and strong commitment of both the patient and, if relevant, the supervisor. The workplace intervention was coordinated by the occupational therapist that mainly concerned feasible solutions for the obstacles for functioning at work. Within the context of this study, two occupational therapists were trained by an expert to coordinate the workplace intervention. More details on the exact content of the intervention can be found elsewhere [17]. All participants received usual rheumatologist-led care as provided in the Netherlands.

Main outcome measures

At-work productivity loss

The primary outcome measure for the cost effectiveness analysis was at-work productivity loss, examined as hours lost from work due to presenteeism per two weeks, measured with the Work Limitations Questionnaire (WLQ) [18]. Presenteeism refers to being present on the job, but being limited in meeting work demands. The WLQ consists of four subscales (time management demands, physical demands, mental-interpersonal demands, and output demands) which are calculated into scores ranging from 0 (no limitations) to 100 (highest limitations). All four subscales have high internal reliability and internal consistency and the WLQ concerning RA shows good validity and reliability [23-25]. The score on all 25 items (which presents the percentage of productivity loss) are multiplied by the number of work hours, resulting in an estimation of presenteeism hours per two weeks.

Quality of Life

The outcome measure for the cost utility analysis was quality adjusted life years (QALY) calculated using EuroQol-5D-5L [26]. The EQ-5D-5L measures health outcomes on five dimensions (mobility, self-care, daily activities, pain/discomfort and

depression/anxiety) ranging from none to major complaints. The Dutch crosswalk value set was used to estimate the utility of health states as described by the respondent [27]. Utilities at baseline and after six months follow-up were used to compute QALYs by means of the area under the curve method [28].

Cost measures and valuation

In health economics, different cost categories can be used. The three categories of costs adhered to in this study were the healthcare costs, the patient- and family costs and the costs in other sectors. Intervention costs were aggregated to estimate the mean total societal costs [29]. Questionnaires with a six month recall period were posted to the respondents at baseline and six months after randomization included health care costs (i.e. health service uptake and medication usage), patient and family costs (e.g. out-of-pocket costs for travel and parking incurred while making use of health services and informal care costs), and costs in other sectors stemming from productivity changes (presenteeism and absenteeism). Costs were indexed to the reference year 2012 (the year in which most inclusions of participants took place). Discounting of costs was not needed because the follow-up period did not exceed one year.

Healthcare costs

Standard Dutch unit prices or average tariffs (when standard prices were not available) according to the Dutch Manual for Costing were used to quantify patient reported healthcare resource utilization into a monetary value [30]. For prescribed medication, prices per dosage were multiplied by their average intakes. Prescription charges were added. The Daily Defined Dosages (DDD) were derived from the Dutch Pharmacotherapeutic Compass [31], the price per dosage was based on drug costs in the Netherlands [32]. Non-prescription drugs were based on their average market prices, including a 6% tax. Medical, personal and professional aids were based on the aid categories as registered by the Dutch care institute [33] or based on their market prices.

Patient and family costs

Patient and family costs comprised informal care costs and travel and parking costs. Informal care costs were calculated by multiplying the hours of informal care with the wages that were valued against the shadow price of the wage rate per hour of a housekeeper [30]. Travel and parking costs were valued by multiplying the distance to the utilized health care service with the average costs per kilometer as presented in the Dutch guideline for costing research [30] and the frequency of used health care services. If applicable, parking costs were added.

Costs in other sectors

Costs in other sectors relate to costs through absenteeism (absent because of sick leave days) and presenteeism (present, but being limited in meeting work demands). The friction costs method was used to calculate absenteeism costs by multiplying the days of production loss till replacement by the average day wage, based on the age- and gender specific productivity levels per paid employee. The time needed to replace and train new employees (i.e. the friction period) contained 92.68 days in the Netherlands in 2012 [34-36]. Presenteeism costs were calculated by correcting for the degree of inefficiency. The inefficiency score for work quantity and –quality, derived from the Productivity and Disease Questionnaire [37], was multiplied with the number of presenteeism days in the previous six months, and multiplied with the wage-specific productivity levels [30].

Intervention costs

Intervention costs were calculated by using a bottom-up approach by multiplying the average time spent on each part with the average wage of the intervention provider. Table 2 presents the intervention costs for the main components.

Data analysis

The analyses were performed according to the intention-to-treat principle. Missing data were replaced by mean imputation using the mean of the outcome per group to replace the missing values per respondent. Parametric analyses (i.e. ANOVA) were conducted to compare the baseline characteristics between both groups. Non-parametric analyses of Kruskal Wallis for continuous variables and Pearson χ^2 tests for categorical variables were used. The 95% confidence intervals around the mean cost differences were attained by non-parametric bootstrap techniques.

For the economic evaluation, an incremental approach was used by calculating the differences between the intervention- and usual care group. The incremental cost effectiveness ratio (ICER) was calculated by dividing the difference in costs (healthcare costs, patient and family costs and intervention costs) by the difference in at-work productivity loss (i.e. hours lost from work due to presenteeism), measured with the WLQ. The incremental cost utility ratio was calculated by dividing the difference in costs (healthcare costs, patient and family costs, costs in other sectors, and intervention costs) by the difference in QALYs. In the cost effectiveness analysis, the costs in other sectors are not included to avoid double counting, because presenteeism serves as the outcome measure (i.e. at-work productivity loss). When conducting the cost utility analysis, the costs in other sectors are included as there is no risk for double counting. A full societal perspective is thereby not applied in the cost effectiveness analysis due to methodological concerns (i.e. double counting)

Both the bootstrapped ICER and ICUR pairs were graphically plotted on a cost effectiveness plane (probabilities in bootstrap on the random sampling with re-

placement based on individual data of the participants). Cost effectiveness acceptability curves were generated if the ICER or ICUR was located in the north-east quadrant (i.e. the intervention produces superior effects at additional costs relative to control) or the south-east quadrant (i.e. the intervention generates superior effects against lower costs).

Risk case- and sensitivity analyses

First, a risk-case analysis, followed by three sensitivity analyses were conducted, aimed to detect whether potential different methodological considerations might impact the results as found in the main analyses.

The risk-case analysis comprised, on an explorative basis, the CEA and CUA for the subsample in which participants at risk were left out. Due to a systematic error in the minimisation procedure, a subgroup of 37 participants was considered at risk to be mistakenly allocated to the control or intervention group. To determine the potential influence of the allocation error on the study results a sensitivity analysis on a subsample in which the 37 participants at risk were left out was performed.

For the sensitivity analyses, first, the human capital theory was used to value absenteeism costs. The human capital approach incorporates the relevant value of the complete lost production time. Although this might produce relatively large estimates of productivity costs, it might be relevant in this chronic ill population, as one could state that lost production time might be longer due to the chronic condition, and thereby not as easy to restore. The total number of sick leave days in the previous six months were multiplied by the gender and age-specific productivity levels per paid employee [30] without accounting for the friction period needed to restore productivity.

Secondly, compensation for lost work was accounted for when calculating absenteeism costs. Absenteeism costs were only calculated when the work was taken over in extra hours or when new personnel were hired to compensate for the production loss. Compensation during normal working hours was assumed not to result in productivity costs.

Finally, costs in other sectors (as recommended when conducting a full economic evaluation from the societal perspective) were accounted for, omitting the risk of double counting for presenteeism costs and effects by only including the costs related to absenteeism.

Results

Participants and baseline characteristics

At baseline, 150 participants were included: 75 were randomized into the intervention group and 75 into the control group. A more detailed description of the patient flow throughout the study is presented in Figure 1.

Figure 1. Flow-chart of patient inclusion

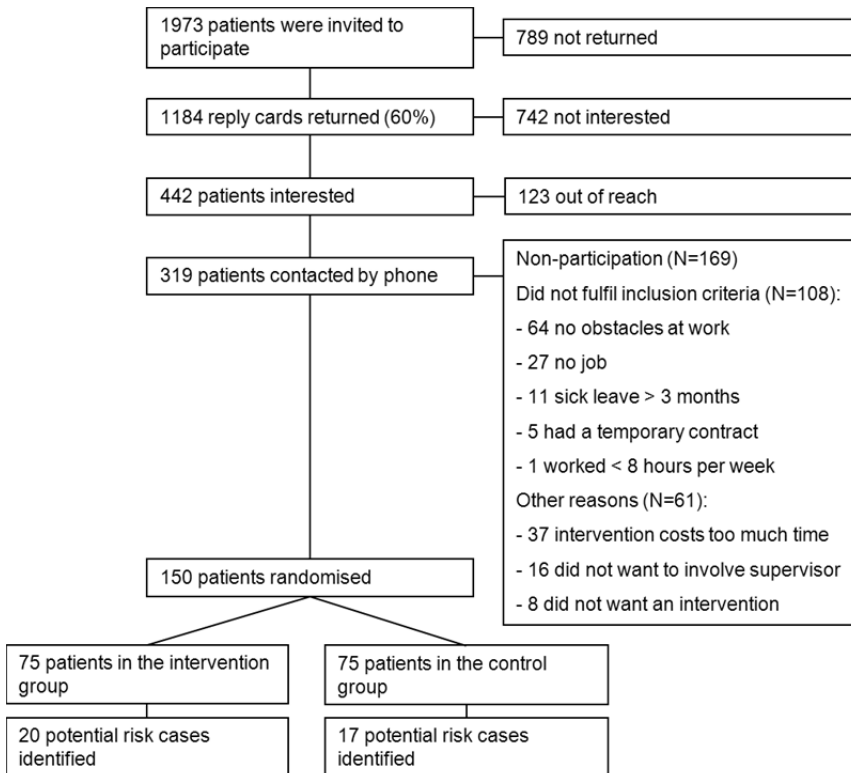


Table 1 presents the baseline demographic characteristics and mean costs for both groups. A significant difference in quality of work and in presenteeism hours (based on the WLQ) was found between both groups. Informal care costs were also significantly different between the intervention- and control group. Managerial positions were more often present in the intervention group (nearly significant) and on average, medical aids were more often used by the control group (near to significance).

IMPROVING WORK PRODUCTIVITY AMONG RA PATIENTS

Table 1. Baseline characteristics (N=150)

	Intervention Group (n=75)	Control Group (n=75)	p-value
Age, mean (sd)	49.75 (8.6)	49.6 (8.7)	0.93 (a)
Female, N (%)	63 (84)	63 (84)	0.604 (b)
Education, N (%) †			
Low	16 (21)	16 (21)	
Intermediate	22 (30)	26 (35)	0.755 (b)
High	37 (49)	33 (44)	
Working hours, mean (sd)	29.22 (10.18)	28.21 (9.9)	0.54 (a)
Managerial position, N (%)	25 (33)	16 (21)	0.064 (c)*
Shift work, N (%)	13 (17)	15 (20)	0.83 (c)
Absenteeism days, mean (sd)	6.14 (15.69)	3.79 (8.91)	0.261 (a)
General Health, mean (sd) #	70.28 (13.8)	73.32 (15.7)	0.236 (a)
Co-morbidity, N (%)	46 (61)	51 (68)	0.125 (c)
Quantity of work, mean (sd) £	8.29 (1.7)	8.74 (1.6)	0.096 (a)
Quality of work, mean (sd) £	8.56 (1.7)	9.17 (1.3)	0.014 (a)**
WLQ based presenteeism hours, mean (sd)	4.55 (2.5)	3.41 (2.79)	0.012 (a)**
Utilities, mean (sd) ††	0.75 (0,14)	0.78 (0,13)	0.112
Costs, mean in € (sd)			
Prescription medication	493.9 (1003.86)	338.75 (506.93)	0.925 (d)
Over-the-counter medicines	40.05 (86.58)	25.58 (58.36)	0.147 (d)
Medical aids	76.61 (209.56)	134.25 (394.48)	0.084 (d)*
Health care service utilization	499.25 (551.77)	585.2 (724.95)	0.901 (d)
<i>Total healthcare costs</i>	<i>1,109.81 (1353.83)</i>	<i>1,083.76 (1088.82)</i>	<i>0.739 (d)</i>
Informal care costs	43.76 (88.95)	23.42 (65.05)	0.048 (d)**
Travel and parking costs	34.66 (35.06)	39.84 (46.42)	0.91(d)
<i>Total patient and family costs</i>	<i>78.42 (100.84)</i>	<i>63.26 (86.17)</i>	<i>0.331 (d)</i>
Absenteeism costs	1,642.39 (4,401.59)	988.56 (2,493.78)	0.74 (d)
Presenteeism costs	1,198.61 (3,028.9)	1,132.31 (4,460.36)	0.4 (d)
<i>Total costs in other sectors</i>	<i>2,841 (5,390.34)</i>	<i>2,120.86 (6,689.75)</i>	<i>0.404 (d)</i>

†Low=preschool, primary school; intermediate= lower and upper secondary; high=tertiary school, university, or postgraduate.
Rated on a visual analogue scale from 0 (bad) to 100 (excellent)
£ Rated on a visual analogue scale from 0 (very bad/low quality or quantity) to 10 (very good/high quality or quantity) | High scores indicate better perceived quality of life outcomes, rated on a scale from (0) bad to 1 (good)
(a) Anova (b) Pearson Chi-squared test (c) Fisher's Exact test (d) Non-parametric Kruskal Wallis test
* Nearly significant| ** Significant at the 5% level

Outcomes

After the intervention, the mean duration of hours lost from work due to presenteeism (i.e. at-work productivity loss) were 4.49 (SD 2.94) in the intervention group and 3.78 (SD 2.39) in the usual care group. The difference in hours lost from work due to presenteeism (i.e. at-work productivity loss) at six months follow-up was not significantly different for both groups (p-value=0.103).

In both groups, the mean utilities after six months follow-up were not significant for both groups ((control group -0.002 (95% CI -0.031 to 0.027) and intervention group -0.011 (95% CI -0.04 to 0.18)). The mean QALY, based on the utility scores at baseline and after six months follow-up, were on average 0.39 (sd 0.054) for the control group and 0.38 (sd 0.053) for the intervention groups, which was not significantly different comparing both groups (p-value=0.09).

Costs

The intervention costs are based on: (i) the integrated care provided by a clinical occupational physician, including three consults, and (ii) the workplace visit and advisory reports done by the occupational therapist. Details on the components and costs can be found in Table 2. Although the intervention costs depend on the uptake of the different intervention components and the type of employment (employee or self-employed) per patient, the overall costs were estimated at €91.84 per participant in the study (Table 3).

Table 2. Intervention costs calculation

Intervention components	Average time spent (hours)	Average wage (€)	Outcome (time*wage) (€)
Care manager - Integrated care by clinical occupational physician (COP)			
First consultation (including: (i) development of treatment plan, (ii) Contact with rheumatologist and patient's OP concerning treatment plan; (iii) Sending communication form to rheumatologist, OT, and patient's OP; and (iv) Facilitate e-mail contact with OT, rheumatologist, and patient's OP)	1.50	29.06	43.59
Second consultation (6 weeks): evaluation with patient	0.33	29.06	9.69
Third consultation (12 weeks): evaluation with patient	0.33	29.06	9.69
Occupational Therapist (OT) - Workplace visit			
Employment contract			
<i>Including: (i) Organizational preparation of the protocol, (ii) Workplace observation , (iii) Inventory 'problems' worker, (iv) Inventory 'problems' supervisor , (v) Conversations on solutions, (vi) Developing advisory reports, (vii) Meeting and contacts with 'COP'</i>	4.33	15.84	68.52
Self-employed			
<i>Including: (i) Organizational preparation of the protocol, (ii) Workplace observation , (iii) Inventory 'problems' worker, (iv) Conversations on solutions, (v) Developing advisory reports, (vi) Meeting and contacts with 'COP'</i>	3.95	15.84	62.57
Occupational Therapist (OT) - Advisory reports			
Instructing solutions, Control / evaluation, Adjusting and continuing	0.74	15.84	11.72
Wage scale Care Manager / Clinical Occupational Physician based on functional description(38): 11 to 14. Chosen median scale 12 (.8)= €5,033 Hourly wage = Wage per month/hours per week/4.33 = 5,033/40/4.33			
Wage scale occupational therapist based on functional description(38): 6 to 10. Chosen median scale 7 (.9)= €2,743 Hourly wage = Wage per month/hours per week/4.33 = 2,743/40/4.33			

The mean healthcare-, patient and family-, intervention-, and productivity costs are presented in Table 3. Regarding healthcare costs, the main cost drivers are medication costs and costs related to hospital care. Significantly higher hospital care costs related to day treatment are found for the control group. Patients in the intervention group used more healthcare provided by the medical specialist since the mean cost difference was significantly higher (Table 3). The mean costs associated with occupational health care (occupational physician and occupational therapist) were highest among the patients in the control group. Absenteeism and presenteeism accounted for the highest costs categories in both groups.

Table 3. Mean component costs

	Mean costs (sd)		
	Intervention group (I)	Control group (C)	Mean cost difference (I-C) (95% CI)¶
Healthcare costs			
General practitioner	46.27 (50.36)	56.07 (55.21)	-10 (-26 to 6)
Medications¥	407 (804.72)	319.50 (539.46)	87 (-120 to 316)
Medical specialist	99.48 (83.06)	78.92 (49.78)	21 (0 to 43)
Occupational physician	3.47 (19.47)	13.34 (41.57)	-10 (-20 to -1)
Occupational therapist	49.16 (232.24)	131.09 (244.83)	-83 (-157 to -2)
Alternative treatment	23.54 (80.17)	39.62 (172.63)	-16 (-62 to 20)
Professional home care	11.46 (48.97)	9.95 (62.01)	1 (-19 to 17)
Psychologist	42.47 (169.29)	69.24 (249.76)	-24 (-94 to 36)
Hospital care			
Day treatment	36.75 (182.49)	144.57 (538.61)	-106 (-254 to -1)
Admission	145.84 (523.540)	67.19 (393.78)	81 (-57 to 232)
Total health care costs	864.95 (1,099.38)	929.5 (1,181.8)	-75 (-433 to 279)
Occupational health care			
Intervention care manager	39.46 (19.98)	n.a.	n.a.
Intervention occupational therapist	52.38 (35.43)	n.a.	n.a.
Total intervention costs	91.84	n.a.	n.a.
Patient and Family costs			
Informal care	30.39 (68.11)	19.24 (50.28)	11 (-7 to 30)
Parking and travel time	39.65 (42.02)	42.75 (47.34)	-3 (-17 to 10)
Total patient and family costs	70.04	61.99	9 (-17 to 35)
Costs in other sectors			
Absenteeism	893.21 (2,120.21)	838.82 (1,959.70)	68 (-589 to 722)
Presenteeism	1,076.28 (2,681.33)	617.31 (1,384.46)	464 (-154 to 1,156)
Total costs in other sectors	1,969.48 (3,915.11)	1,456.13 (2,440.13)	518 (-494 to 1,632)

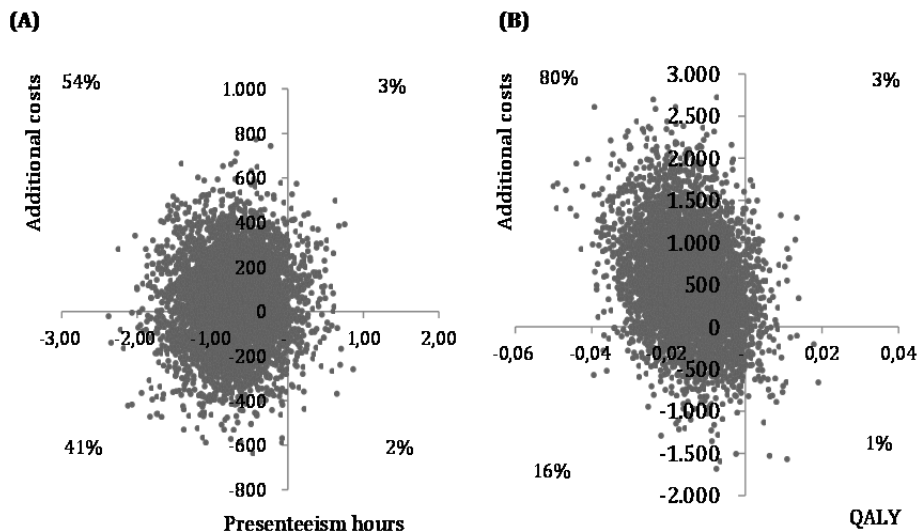
¶Mean cost differences and 95% confidence intervals around the mean cost differences were attained by non-parametric bootstrap techniques| ¥Including both prescription and over-the-counter medications| n.a.=not applicable

Economic Evaluation

The main cost effectiveness and cost utility analyses show ICERs in similar directions: more expensive and less effective (respectively 49.03 and -37,073.45) (Table 4). Note that the ICER for at-work productivity loss was positive because higher scores indicate more presenteeism hours, and thus worse outcomes when compared with control. In both cost effectiveness planes, the majority of the bootstrapped cost-effect pairs (ICER 54%; ICUR 80%) are located in the northwest

quadrant (Figure 2), also indicating that the intervention was more expensive than the control in obtaining an additional effect.

Figure 2. Costs effectiveness planes for the incremental cost effectiveness ratio (A) and the incremental cost utility ratio (B)



Risk- case and sensitivity analyses

The results of the sensitivity analyses can be found in the lower part of Table 4. The overall conclusions did not change when solely analysing the subgroup of participants not at risk of being mistakenly randomized to one of the groups (n=37). The results of the sensitivity analyses testing the robustness of the cost utility analysis results (using the results from the human capital theory used to value absenteeism costs, and compensation for lost work when calculating absenteeism costs) did not differ from the main analysis. Results of the sensitivity analysis in which the costs of absenteeism were included and the effect in productivity (number of hours lost from work due to presenteeism) were estimated did neither change the overall conclusions. In all sensitivity analyses the majority of the incremental cost-effect pairs remained located in the North West quadrant, indicating that the intervention was less effective and more costly when compared with control.

Table 4. Total mean effects and costs during 6-months follow-up, incremental cost effectiveness ratios and 95% confidence interval, distribution of the cost-effect pairs on the cost effectiveness plane

		Sample Size		Distribution CE Plane (%)							
I	C	Effect	Δ Costs (€)	Δ Effect	ICER (Δ Costs / Δ Effects)	ICER 95% CI	NE	SE	SW	NW	
Main Analyses											
Cost effectiveness	75	75	Presenteeism	35.33	0.72	49.03	-106,854 to 1,179	3	2	41	54
Cost utility	75	75	QALY [¶]	548.69	-0.01	-37,073.45	-216,987 to 190,100	3	1	16	80
Risk-case Analyses											
Cost effectiveness: without potential risk cases	55	58	Presenteeism	22.71	0.65	34.97	-3,171 to 2,836	6	4	41	48
Cost utility: without potential risk cases	55	58	QALY [¶]	195.09	-0.02	-12,119.61	-233,466 to 239,670	2	4	36	58
Sensitivity Analyses											
Costs utility: human capital approach	75	75	QALY [¶]	513.35	-0.01	-34,686.06	-268,342 to 158,805	3	1	15	81
Cost utility: compensation productivity loss	75	75	QALY [¶]	555.74	-0.01	-37,549.79	-282,275 to 174,133	4	1	8	88
Cost effectiveness: absenteeism costs included	75	75	Presenteeism	89.72	0.72	124.50	-3,521 to 3,250	3	2	39	56
[¶] hours lost from work due to presenteeism, measured with the Work Limitations Questionnaire (WLQ) [‡] quality adjusted life years calculated using EuroQol-5D-5L NE = North East quadrant indicating the intervention is more effective and more costly compared with control SE = South East quadrant indicating the intervention is more effective and less costly compared with control SW = South West quadrant indicating the intervention is less effective and less costly compared with control NW = North West quadrant indicating the intervention is less effective and more costly compared with control.											

Discussion

This study presents the cost effectiveness and cost utility analysis of an integrated care program for workers with rheumatoid arthritis to improve their work productivity. The intervention cannot be regarded cost effective. Although not significant, both costs and effects were more favourable in the control group than in the intervention group over the measurement period of six months follow-up.

A previously conducted study evaluating the effectiveness of an integrated care program for patients with chronic low back pain found that integrated care was effective on return to work (HR 1.9), and sustainable return to work ($P=0.003$) after a measurement period of 12 months [39]. The same integrated care program was even found cost effective when compared with usual care for return to work and QALYs gained with 12 months' follow-up [11]. The probability that the intervention as described in this study might become cost effective on a longer time span might be possible. However, a study evaluating the cost effectiveness of integrated, multidisciplinary care as compared with usual care for patients with moderate to severe chronic hand dermatitis was neither cost effective nor effective after 12 months follow-up, in contrast to the findings after six months follow-up [40].

Before outweighing the costs and benefits of the integrated care- and participatory workplace intervention under study, a few remarks of the study should be noted. Firstly, the study was performed alongside a randomized controlled trial. Although a subgroup of the participants risked being mistakenly allocated to one of the groups, the results were reasonably robust. This RCT design allowed for the economic evaluation to be conducted in a real life setting collecting relevant cost and effect data. However, due to the character of the intervention, patients, therapists and researchers could not be blinded for the allocated treatment after randomization which might increase the risk of bias. Furthermore, the generalizability of the results to other work-disabled populations is unknown as the intervention program was tailored to workers with rheumatoid arthritis.

Secondly, this study reflected the societal perspective by including all costs regardless of who uses the resources or who benefits from them. A key note here might be the fact that the chosen perspective might influence the results. However, when not accounting for the societal perspective, the costs are merely shifted, not really saved. One might argue in the context of this study the insurer's or healthcare perspective might have been interesting to incorporate in the analyses. An insurer might be looking to reduce its expenditure for medical costs. However, in the process the insurer simply shifts the costs from itself to someone else. For example, outpatient care may reduce the medical expenses for a procedure of in-hospital care. However, the costs of the patients did not disappear; they shifted from the insurer to the patient (e.g. when informal care has to be provided and the informal caregiver needs to take time off from work). The same would hold for the productivity losses, shifting the costs of lost productivity from the insurer to the company.

Although the societal perspective is the recommended option, and therefore applied in this study, it might also be helpful to describe the costs from other perspectives, for example by describing in detail what the intervention costs were (comparable to the information provided in Table 2) and to demonstrate the return on investment or the value of providing the intervention for companies and employers. Other perspectives can therefore also be useful to demonstrate the value in supporting the intervention's aims.

Thirdly, this study presents the collection of occupational health care costs (i.e. intervention costs) in a detailed manner. By transparently monetizing the intervention, decision-makers and stakeholders with different perspectives are provided with comprehensive intervention costs to include in future (economic) evaluations. A significant difference in occupational health care costs was found, which is probably attributable to the fact that the intervention group receives this care in the intervention. In the control group, occupational health care was contracted via the regular path and not by the uptake of the intervention components. Therefore, costs related to contacts with the care manager, occupational physician, workplace visits and advisory reports were not applicable since these interventions were not offered to this group, explaining the significant differences.

Fourthly, the lack of effectiveness might be explained by the baseline values being already good, leaving less room for improvement (especially for quality of life), or the measurement instrument might not have been sensitive enough to detect a difference (especially for quality of life) for this disease-specific population. Furthermore, the nature of the intervention is promoting (work productivity) and preventing (work disability) whereby the effects are potentially long-term, requiring follow-ups exceeding six months. The time for follow-up in this study might therefore have been too short to detect effectiveness as long-term follow-up is essential to evaluate the effects of a new innovative intervention critically. Another potential explanation for the lack of effectiveness, especially regarding presenteeism hours (i.e. at-work productivity loss), can be found in the so-called phenomena 'response-shift'. Hereby, it is hypothesized that patients who are confronted with health-related work disability, by for example actively paying attention to presenteeism as this is being 'intervened', are faced with the necessity to change their internal standards, values and conceptualization. This might result in more presenteeism hours, because of the attention that is being paid to it. Although response shift is routed in health-related quality of life research, it refers to a change in the meaning of one's self-evaluation and can therefore occur in any field where self-reports are required [41].

Finally, the cost data are highly skewed and common standards regarding the inclusion, identification, measurement and valuation of productivity costs, especially absenteeism- and presenteeism costs, are lacking [14, 42, 43]. However, despite of the different methods used in the sensitivity analyses to provide crude estimates of absenteeism- and presenteeism costs associated with the intervention, no differences in cost-effect ratios were detected. Furthermore, this study is about the joint

distribution of differences in costs and effects, whereby neither cost nor effect differences should be individually significant to show clear cost effectiveness [44].

Conclusion

Although cautiously, it can be concluded that the intervention was not cost effective after six months follow-up. At the moment, no evidence to support its implementation on cost-related grounds is provided.

References

1. Verstappen SM, Boonen A, Bijlsma JW, Buskens E, Verkleij H, Schenk Y, et al. Working status among Dutch patients with rheumatoid arthritis: work disability and working conditions. *Rheumatology (Oxford)*. 2005 Feb;44(2):202-6.
2. Zirkzee EJ, Sneep AC, de Buck PD, Allaart CF, Peeters AJ, Runday HK, et al. Sick leave and work disability in patients with early arthritis. *Clin Rheumatol*. 2008 Jan;27(1):11-9.
3. Burton W, Morrison A, Maclean R, Ruderman E. Systematic review of studies of productivity loss due to rheumatoid arthritis. *Occup Med (Lond)*. 2006 Jan;56(1):18-27.
4. Boonen A, Severens JL. The burden of illness of rheumatoid arthritis. *Clin Rheumatol*. 2011 Mar;30 Suppl 1:S3-8.
5. Goetzel RZ, Long SR, Ozminkowski RJ, Hawkins K, Wang S, Lynch W. Health, absence, disability, and presenteeism cost estimates of certain physical and mental health conditions affecting U.S. employers. *J Occup Environ Med*. 2004 Apr;46(4):398-412.
6. Li X, Gignac MA, Anis AH. The indirect costs of arthritis resulting from unemployment, reduced performance, and occupational changes while at work. *Med Care*. 2006 Apr;44(4):304-10.
7. Lundkvist J, Kastang F, Kobelt G. The burden of rheumatoid arthritis and access to treatment: health burden and costs. *Eur J Health Econ*. 2008 Jan;8 Suppl 2:S49-60.
8. Wolfe F, Allaire S, Michaud K. The prevalence and incidence of work disability in rheumatoid arthritis, and the effect of anti-tumor necrosis factor on work disability. *J Rheumatol*. 2007 Nov;34(11):2211-7.
9. Krol M, Brouwer W, Rutten F. Productivity costs in economic evaluations: past, present, future. *Pharmacoeconomics*. 2013 Jul;31(7):537-49.
10. LHV-NVAB-Leidraad voor huisarts en bedrijfsarts bij de sociaal-medische begeleiding bij arbeidsverzuim. *De Huisarts/Tijdschr Bedrijfsgeneesk Verzekeringsgeneesk*. 2002;S1(19).
11. Lambeek LC, Bosmans JE, Van Royen BJ, Van Tulder MW, Van Mechelen W, Anema JR. Effect of integrated care for sick listed patients with chronic low back pain: economic evaluation alongside a randomised controlled trial. *BMJ*. 2010;341:c6414.
12. Badamgarav E, Croft JD, Jr., Hohlbauch A, Louie JS, O'Dell J, Ofman JJ, et al. Effects of disease management programs on functional status of patients with rheumatoid arthritis. *Arthritis Rheum*. 2003 Jun 15;49(3):377-87.
13. Vermeulen SJ, Heymans MW, Anema JR, Schellart AJ, van Mechelen W, van der Beek AJ. Economic evaluation of a participatory return-to-work intervention for temporary agency and unemployed workers sick-listed due to musculoskeletal disorders. *Scand J Work Environ Health*. 2013 Jan;39(1):46-56.
14. Krol M, Brouwer W, Rutten F. Productivity Costs in Economic Evaluations: Past, Present, Future. *Pharmacoeconomics*. 2013 Apr 26(7):537-49.
15. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *Pharmacoeconomics*. 2013 May;31(5):361-7.
16. Drummond M, Manca A, Sculpher M. Increasing the generalizability of economic evaluations: recommendations for the design, analysis, and reporting of studies. *Int J Technol Assess Health Care*. 2005 Spring;21(2):165-71.
17. van Vilsteren M, Boot CR, Steenbeek R, van Schaardenburg D, Voskuyl AE, Anema JR. An intervention program with the aim to improve and maintain work productivity for workers with rheumatoid arthritis: design of a randomized controlled trial and cost effectiveness study. *BMC Public Health*. 2012;12:496.
18. Lerner D, Amick BC, 3rd, Rogers WH, Malspeis S, Bungay K, Cynn D. The Work Limitations Questionnaire. *Med Care*. 2001 Jan;39(1):72-85.
19. de Zwart BC, Broersen JP, van der Beek AJ, Frings-Dresen MH, van Dijk FJ. Occupational classification according to work demands: an evaluation study. *Int J Occup Med Environ Health*. 1997;10(3):283-95.

20. Evans S, Royston P, Day S. Minim: allocation by minimisation in clinical trials 2011 [2011 July 6]. Available from: <http://www-users.york.ac.uk/~mb55/guide/minim.htm>.
21. Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials. a review. *Control Clin Trials*. 2002 Dec;23(6):662-74.
22. Altman DG, Bland JM. Treatment allocation by minimisation. *BMJ*. 2005 Apr 9;330(7495):843.
23. Walker N, Michaud K, Wolfe F. Work limitations among working persons with rheumatoid arthritis: results, reliability, and validity of the work limitations questionnaire in 836 patients. *The Journal of rheumatology*. 2005 Jun;32(6):1006-12.
24. Lerner D, Reed JI, Massarotti E, Wester LM, Burke TA. The Work Limitations Questionnaire's validity and reliability among patients with osteoarthritis. *Journal of clinical epidemiology*. 2002 Feb;55(2):197-208.
25. Allaire S. Measures of adult work disability *Arthritis & Rheumatism* 2003;49(5S):85-9.
26. EQ-5D User Guide. Rotterdam Erasmus Univeriteit Rotterdam, centrum voor gezondheidsbeleid en recht, 1995.
27. EuroQol. EQ-5D-5L Value Sets [cited 2014]. Available from: <http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-value-sets.html>.
28. Matthews JN, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. *BMJ*. 1990 Jan 27;300(6719):230-5.
29. Drummond MF. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford; 2005.
30. Hakkaart L, Tan SS, Bouwmans, CAM. Manual for cost research. Methods and standards costs for economic evaluations in healthcare [in Dutch: Handleiding voor kostenonderzoek. Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg]. Rotterdam, the Netherlands: Instituut voor Medical Technology Assessment, Erasmus Universiteit Rotterdam, 2010.
31. Healthcare institute Netherlands. Pharmacotherapeutic compass [in Dutch: College van Zorgverzekeringen. Farmacotheurapeutisch kompas]. [Cited September 2012]; Available from: <http://www.fk.cvz.nl/>.
32. Zorginstituut Nederland. Drug costs in the Netherlands 2014 [cited 2014 August 2014]. Available from: <http://www.medicijnkosten.nl/>.
33. Zorginstituut Nederland. Medical aids categories 2014 [cited 2014 August 2014]. Available from: www.gipdatabank.nl.
34. Koopmanschap MA, van Ineveld BM. Towards a new approach for estimating indirect costs of disease. *Soc Sci Med*. 1992 May;34(9):1005-10.
35. Oostenbrink JB, Koopmanschap MA, Rutten FF. Standardisation of costs: the Dutch Manual for Costing in economic evaluations. *Pharmacoeconomics*. 2002;20(7):443-54.
36. Statistics Netherlands. Jobs (outstanding, new and filled) [cited 2014]. Available from: <http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=80857NED&LA=NL>.
37. Koopmanschap MA. PRODISQ: a modular questionnaire on productivity and disease for economic evaluation studies. *Expert Rev Pharmacoecon Outcomes Res*. 2005 Feb;5(1):23-8.
38. CAO Universitaire medisch centra 2014. Available from: <http://www.nfu.nl/publicaties/cao-universitair-medische-centra/print-cao>.
39. Lambeek LC, van Mechelen W, Knol DL, Loisel P, Anema JR. Randomised controlled trial of integrated care to reduce disability from chronic low back pain in working and private life. *BMJ*. 2010;340:c1035.
40. van Gils RF, Bosmans JE, Boot CR, Rustemeyer T, van Mechelen W, van der Valk PG, et al. Economic evaluation of an integrated care programme for patients with hand dermatitis. *Contact dermatitis*. 2013 Sep;69(3):144-52.
41. Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. *Soc Sci Med*. 1999 Jun;48(11):1507-15.
42. Prasad M, Wahlqvist P, Shikiar R, Shih YC. A review of self-report instruments measuring health-related work productivity: a patient-reported outcomes perspective. *Pharmacoeconomics*. 2004;22(4):225-44.

CHAPTER 9

43. Noben CY, Evers SM, Nijhuis FJ, de Rijk AE. Quality appraisal of generic self-reported instruments measuring health-related productivity changes: a systematic review. *BMC Public Health*. 2014;14:115.
44. Petrou S, Gray A. Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. *BMJ*. 2011;342:d1548.

CHAPTER 10

Short-term cost effectiveness and cost utility of a training program to increase GP awareness to work-related problems: A randomized controlled trial*

Cindy Noben
Cornelis de Kock
Antoine Lagro-Janssen
André Knottnerus
Angelique de Rijk
Frans Nijhuis
Romy Steenbeek
Silvia Evers

** the included economic evaluation results are based on interim-analyses as data collection is ongoing*

Abstract

Objectives: Although diverse initiatives have been taken to increase awareness on work-related health issues, general practitioner's (GP) guidelines (at least in the Netherlands) do not yet include advice on timely addressing work-related problems. Therefore, a tailored educational training program for GPs was developed to increase GP awareness to work-related health problems aiming for better treatment strategies to benefit both individual patients and society as a whole. The aim of the study was to determine whether GP educational training was preferable to usual GP care in terms of work-related self-efficacy for employees and in terms of employees' quality of life.

Methods: An economic evaluation from a societal perspective was conducted alongside a cluster randomized controlled trial. Patients between 18-63 years of age in paid work for at least twelve hours per week were eligible. The primary outcomes were work-related self-efficacy, measured by the return-to-work self-efficacy (RTW-SE) scale and quality adjusted life years (QALYs) based on the euroqol-5D-5L. Cost data were collected via questionnaires directly after the consult and six months later. Data were analyzed according to the intention-to-treat principle. Missing data were imputed using mean imputation techniques. Bootstrapped cost effect pairs were plotted on the cost effectiveness planes.

Results: The difference in RTW-SE was minimal between the intervention- and control group. A negative incremental cost effectiveness ratio (ICER) (-€ 1,858.98) was found, meaning that the intervention yielded less effects albeit higher costs. The cost utility analysis resulted in similar findings: the intervention was less effective and more costly than the control condition. The gender sensitivity analyses all confirmed these findings. Based on the sensitivity analyses of the subgroup employees with a permanent employment contract and the one accounting for baseline differences at the start of the study, the intervention resulted in less effects albeit fewer costs when compared to the control condition.

Conclusion: The training program to increase GPs' awareness to work-related problems was not cost effective from the societal perspective regarding both RTW-SE and QALY outcomes. Based on cost-related grounds the intervention could therefore not be recommended for implementation.

Objective

Although most health professionals will acknowledge the importance of the patient's work and workability, this is often not reflected in the attention they pay to these issues when treating patients. Even general practitioners (GPs), who are trained to take the background and context of their patients into consideration, do not always pay attention to work-related health issues [1, 2]. An important reason for the lack of attention for work-related health issues among Dutch GPs may be the fact that they, unlike their colleagues in most other countries, do not have a formal role in sickness certification or in the rehabilitation to work of disabled workers [3, 4]. Furthermore, most GPs appear not to consider sickness certification to be amongst their most meaningful activities [5, 6]. Nonetheless, a more active role of primary care providers with respect to their patients' working context is expected to be beneficial because of the (two way) relation that exists between health and work. Previous studies have demonstrated the relevance of work-related factors in both care- and cure settings [7] and recent Dutch research on GPs and work has demonstrated the importance of collaboration between general and occupational practice [8]. However, collaboration remains poor, despite several measures being taken to improve this interaction [8]. In 2005 the Health Council of the Netherlands published a set of guidelines about the management of diseases that frequently lead to long term or permanent absenteeism to improve work awareness in curative care professionals [9]. More recently, the Dutch College of General Practitioners published its 'core values'. In this publication, the importance of work as one of the contextual factors which GPs should structurally and proactively be paying attention to, is specifically emphasized [10]. Although a number of multidisciplinary- and GP guidelines have been developed and revised to include recommendations on occupational health, most GP guidelines do not yet include advice on timely addressing work-related problems.

Because male and female doctors act differently and discuss different subjects during consultations when confronted with similar problems and because male and female patients differ in consultation behavior, this study also addresses gender aspects in relation to work- and health related problems. Previous research found that gender influences doctor-patient communication and that sick-listing was shown to be influenced by gender [11-14]. To increase GP awareness on work-related health problems and to improve treatment strategies, taking into account gender aspects related to work and health, a tailored educational GP training program was developed with the expectation that this intervention would benefit both individual patients and society as a whole by decreasing healthcare costs.

This study assessed the cost effectiveness and cost utility of the GP educational training program. The aims of the study were (i) to determine whether the GP educational training program was cost effective when compared to usual GP care for employees regarding return to work self-efficacy (RTW-SE); and (ii) to detect

whether the GP educational training program was preferable in terms of employees' quality of life when compared to usual GP care.

Methods

The economic evaluation was conducted alongside a cluster randomized controlled trial. Extensive information regarding the methodological details of this trial has been previously published [15]. Randomization was performed at practice level in order to prevent contamination. A block randomization scheme using a block size of two was generated to assign GPs to either the educational training program or the control group (1:1 allocation ratio). Patients who visited GPs who completed the GP educational training were compared with patients visiting their GP in the usual care setting. Patients were approached by the receptionist while in the waiting room with a short questionnaire to establish eligibility to participate and, if so, obtain their consent. If patients consented, the first questionnaire was handed to them directly after the consultation, followed by a second after six months. The GP's role in data collection was kept as limited as possible to minimize contamination of the care provision in the GP control group.

Study Population

Patients were included if they: (i) were 18-63 years of age, (ii) had paid work for at least twelve hours per week, (iii) sufficiently understood the Dutch language to fill in a questionnaire. The upper age limit of 63 was chosen to lower the chance that patients would drop out of the questionnaire study during follow-up due to retirement.

Educational training intervention

GPs in the intervention group received an accredited five hour training intervention based on the findings of previously conducted focus groups [16]. During the training, lectures and interactive workshops were brought whereby the following seven items were covered: (i) the societal relevance of the connection between work and health and the potential role for primary healthcare; (ii) reflection by the participants' on their own 'usual care' for work-related problems, including awareness of gender differences among patients and care providers in communication styles; (iii) the legislation regarding work and absenteeism, the role of the occupational physicians (OPs) and how to achieve an effective collaboration between GPs and OPs; (iv) knowledge exchange on gender aspects influencing occupation and work-related problems; (v) methods to assess patients and instructions on how to empower them to solve their work-related problems (vi) advice regarding the struc-

tural registration of work-related data (including occupation, number of hours working per week, possible occupational relation of the presented health problem, and adequate coding of work-related problems); and (vii) instructions on the study logistics and data collection. Two months after the initial training a three hour booster training event took place [15].

Care as usual

The control group received care as usual. Receptionists, who were instructed and supported by a research assistant when possible, were responsible for the data collection in order to reduce the possibility that this process influenced the provision of usual care. After the data collection was completed, the GP control group was offered the same training as the intervention group as an incentive for adherence to the study protocol [15].

Outcome measures

The primary outcome for the cost effectiveness analysis was work-related self-efficacy, measured by the eleven item Return-to-Work Self-Efficacy scale (RTW-SE). The RTW-SE scale measures the extent to which people feel able to handle the demands of their job which can be rated from '1' (completely disagree) to '6' (completely agree)[17]. A mean score of the eleven items at follow-up was used to compute changes in the RTW-scale score and differences between both groups. The internal consistency (Cronbach's alpha) of the scale is excellent ($>.8$) [17].

The main outcome for the cost utility analysis was quality adjusted life years (QALY), measured by using the EuroQoL five dimensions, five levels (EQ-5D-5L). These five dimensions are: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. To estimate the utility of health states described by the patients the EQ-5D-5L crosswalk value set with the Dutch tariff was used [18]. Then, based on the utilities, QALYs were calculated using the 'area under the curve' method [19]. Higher QALYs indicate more improvement in quality of life.

Resource use measurement and valuation

Cost data were collected from a societal perspective via questionnaires directly after the consultation and six months later. All costs were adjusted to the index year 2012 using consumer price indices [20]. Discounting of costs was not necessary because the cost measures did not exceed one year. Four main cost categories were used (intervention-, healthcare-, patient- and family-, and productivity costs). The intervention costs were calculated using a bottom-up approach (see Appendix 1).

Health care costs, covering care provider resource utilization (i.e. visits for primary and secondary care), home care, alternative care and medication, were calculated according to the guidelines for health care in the Netherlands by using

the Dutch manual for costing [21]. Tariffs or an average price from providers were used when standard cost prices were not available. The costs of prescription medicines were calculated based on the price per dosage for medication costs in the Netherlands [22, 23]. Over the counter medication prices were based on market prices (including 6% VAT) and medical and personal aids (e.g. orthotics, mobility aids, etc.) were based on costs per user within the aid category provided by the Dutch care institute [24].

Patient and family costs were incorporated by assessing informal care costs and travel and parking costs incurred while utilizing health care. Informal care was valued against the 'shadow price' (i.e. the price used when an official price is not available) of the wage rate per hour of a housekeeper. Travel costs were calculated by multiplying average distance with the standard price weights as derived from the Dutch guideline for costing research [21].

Patient's productivity was measured using the productivity and disease questionnaire (PRODISQ) [25]. The PRODISQ measures productivity loss from paid work (i.e. both absenteeism and presenteeism) in terms of total sick leave. Absenteeism refers to the total number of days lost from work and presenteeism refers to reduced work quantity and quality while at work [25]. Both absenteeism days and the inefficiency score were multiplied by the mean age- and gender wage rates of the Dutch population [21]. Productivity cost calculations were based on the friction cost approach which states that sick employees are replaced after a certain period of time called the friction period. After this friction period, no more lost productivity costs are assumed [26]. The friction period of 92.68 days for the Netherlands was estimated based on the level of unemployment in 2012 [27].

Analysis

Data were analyzed according to the intention-to-treat principle. Missing data were handled using SPSS missing value analysis on item level (mean imputation). Baseline characteristics were compared for both groups by using analysis of variances (Anova), non-parametric Kruskal-Wallis analysis for continuous variables or Pearson χ^2 test for categorical comparisons of proportions.

Because of the highly skewed cost distribution, bootstrapping with 1000 replications was used to estimate bootstrap confidence intervals around the cost differences. The incremental cost effectiveness ratio (ICER) was calculated by dividing the incremental costs by the incremental effects. By dividing the incremental costs by the differences in QALYs, the incremental cost utility ratio was calculated. Non-parametric bootstrapping was also used to estimate the uncertainty surrounding the incremental cost effectiveness ratio and the incremental cost utility ratio (5,000 replications). Bootstrapped cost effect pairs were plotted on the cost effectiveness planes. Cost effectiveness acceptability curves demonstrate the probability of the intervention being cost effective at a certain threshold, i.e. the amount of money society is willing to pay to gain one unit of effect (e.g. €18,000 per QALY). The will-

ingness-to-pay (WTP) in the field of occupational health care is an unknown quantity. We therefore used a series of WTP thresholds and calculated the probability that the intervention is more acceptable than usual care from a cost effectiveness point of view for each of the WTP thresholds (unless >70% of the bootstrapped ICERs are located in the North-west quadrant of the plane, indicating higher costs and fewer effects and the intervention is obviously dominated by the control condition).

Sensitivity analyses

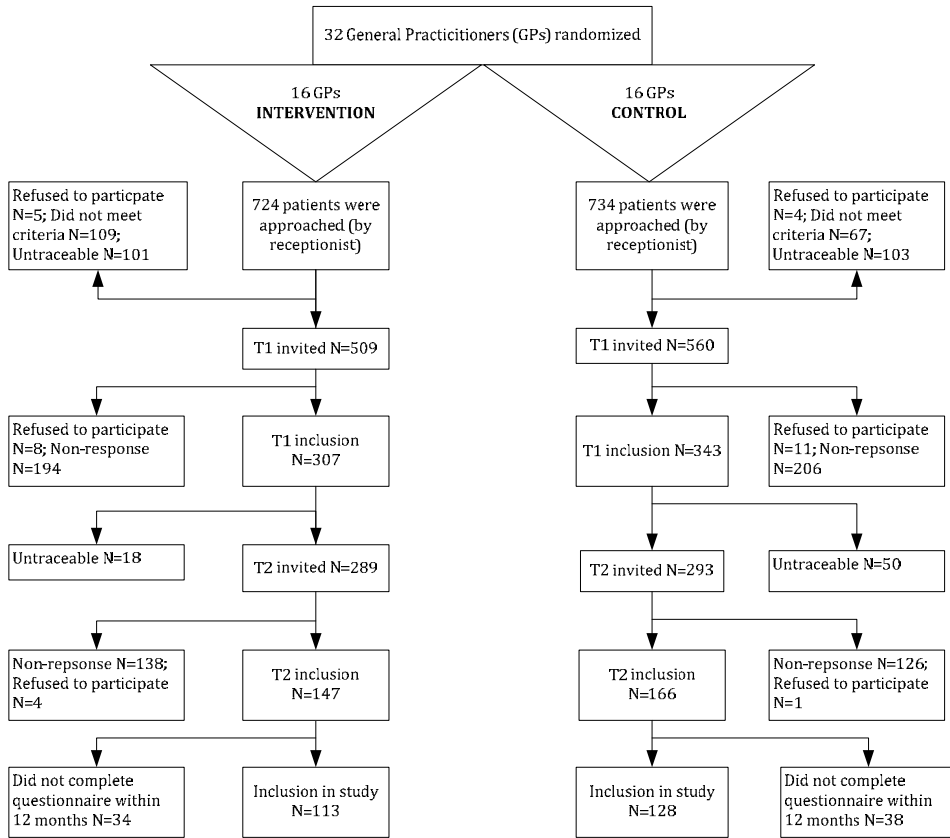
Six sensitivity analyses were conducted to assess the robustness of the findings. First, four subgroup analyses were conducted to test for participant heterogeneity by assessing gender differences among patients and GPs. Then, because a significant (Pearson Chi-square test significant at the 5% level) difference in job contract was found between the intervention and control group, the economic evaluation was redone including only the subgroup employees with a permanent employment contract. Because of overall lower self-rated health and more absenteeism days in the intervention group at the start of the study, the final sensitivity analysis was performed by accounting for the individual differences between patient baseline and the subsequent measurements (also called delta-adjustment) [28].

Results

Sample

An overview of the inclusion procedure is presented in Figure 1. Between February 2012 and January 2013 the GP's receptionists approached patients in the waiting room with a short questionnaire to deploy the selection criteria. When meeting the selection criteria patients could be included; they signed informed consent and then received the first questionnaire. A second questionnaire followed after six months, with a maximum completion time of twelve months. In the end 241 patients were included in the study; 113 in the intervention group and 128 in the control group. As can be seen in Figure 1 many patients were marked 'untraceable' at the start of the study and after completing the first questionnaire because no personal details were available at the GP practice or patients did not provide their contact details to the research team and could therefore not be directly approached.

Figure 1. Inclusion procedure



The population characteristics are presented in Table 1. A small statistically significant difference (p-value 0.045 Pearson Chi-square) was found for job contracts. Most employees were in permanent employment in both groups. The average health care utilization costs were highest in the control group. All other cost categories were highest for the intervention group. At baseline, absenteeism costs accounted for the largest part of the costs in both groups (see Table 1).

INCREASING GP AWARENESS TO WORK-RELATED PROBLEMS

Table 1. Baseline characteristics (N=241)

	Intervention (n=113)	Control (n=128)	p-value
Age, mean (sd)	45.55 (9.8)	47.23 (10.2)	0.19 (a)
Gender, N (%)			
Female	70 (62)	72 (56)	0.37 (b)
Male	43 (38)	56 (44)	
Education, N (%)†			
Low	4 (4)	2 (2)	0.61 (b)
Intermediate	73 (65)	83 (65)	
High	36 (31)	43 (33)	
Working hours, mean (sd)	29.74 (10.7)	30.94 (11.3)	0.40 (a)
Job contract, N (%)			
Entrepreneur	3 (3)	13 (10)	0.045*(b)
Employee (permanent employment)	91 (81)	102 (80)	
Employee (temporary employment towards permanent employment)	6 (5)	5 (4)	
Employee (temporary employment/fixed term)	9 (7)	7 (5)	
Interim / temporary worker	0	1 (1)	
Social Employment Law worker	4 (4)	0	
Supervisor, N (%)			
One	103 (91)	109 (85)	0.15 (b)
More than one	10 (9)	19 (15)	
Managerial position, N (%)	35 (31)	41 (32)	0.86 (b)
Shift work, N (%)	20 (18)	25 (20)	0.51 (b)
Job sector, N (%)			
Craft and industry	12 (11)	17 (13)	0.64 (b)
Transport	3 (3)	5 (4)	
Administration	16 (14)	12 (9)	
Commercial	10 (9)	13 (10)	
Services	5 (4)	10 (8)	
Health and care	22 (19)	29 (23)	
Teaching	12 (11)	8 (7)	
Specialist (discipline, expertise)	3 (3)	7 (5)	
Agriculture	2 (2)	3 (2)	
Other	28 (24)	24 (19)	
Self-rated general health‡, mean (sd)	73.3 (22)	75.7 (21.3)	0.41 (a)
Long term disability, N (%)			
No	67 (60)	78 (61)	0.9 (b)
Yes	45 (40)	50 (39)	
Days absent from work past 6 months, mean (sd)	9.32 (23)	6.49 (21)	0.37 (a)
EQ-5D-5L Utilities, mean (sd)‡	0.84 (0.1)	0.84 (0.1)	0.97 (a)
RTW-Self efficacy, mean (sd)	4.71 (1.3)	4.83 (1.1)	0.43 (a)
Costs, mean in € (sd)			
Health care utilization	438.41 (580)	530.08 (207)	0.12 (c)
Medication and aids	139.37 (317)	106.27 (216)	0.48 (c)

Patient and family	52.57 (104.5)	36.07 (70.2)	0.13 (c)
Informal care	12.78 (74.7)	3.52 (14.6)	0.24 (c)
Travel and parking	39.79 (46.51)	32.55 (61.4)	0.17 (c)
Productivity	2,332.89 (4,937)	1,569.94 (4,322.1)	0.32 (c)
Absenteeism	2,282.67(4,906.4)	1,523.53 (4,290.7)	0.41 (c)
Presenteeism	50.22 (64.24)	46.41 (66.7)	0.33 (c)

†Low=pre- or primary school; intermediate=lower- or upper secondary; high=tertiary school, university, or postgraduate

¥ On a rating scale from 0 (worse health) to 100 (perfect health)

‡High scores indicate better perceived health outcomes

(a)Anova (b)Pearson Chi-squared (c)Non-parametric Kruskal Wallis

* Significant at the 5% level

Main outcomes

The mean utilities (0.84) at baseline were similar and did not differ significantly ($p > 0.05$) between the intervention and the control group. The mean utilities improved slightly after six-month follow-up (control group 0.86 and intervention group 0.85). The mean QALYs in the intervention group (0.42; SD 0.06) and the control group (0.42; SD 0.07) were comparable. The mean RTW-SE score at six months follow-up improved in the control group (5.03 (SD 0.96)) and remained stable (4.8 (SD 1.2)) in the intervention group. The incremental difference in effects for QALYs and RTW-SE are presented in Table 3.

Cost analysis

The intervention costs were almost negligible (€0.48) (see Table 2). Furthermore, Table 2 shows the healthcare costs at six month follow-up, which were higher in the intervention group (€516) than in the control group (€426). The mean cost difference (almost €90) was in favor of the control group. The average healthcare utilization costs are higher in the intervention group, whilst at the start of the study healthcare utilization was highest in the control group. Medication and aids costs remained relatively stable over time. Incremental family and patient costs (€7) were in favor of the intervention group. The patient and family costs in the control group remained stable over time; a decrease within these costs occurred in the intervention group six months after inclusion. The productivity costs (mean €2,211.79 and €2,660.92 respectively in control and intervention group) continued to account for the largest part of the mean total costs. Productivity costs increased in both groups after six months follow-up and the incremental difference (€449) was in favor of the control group.

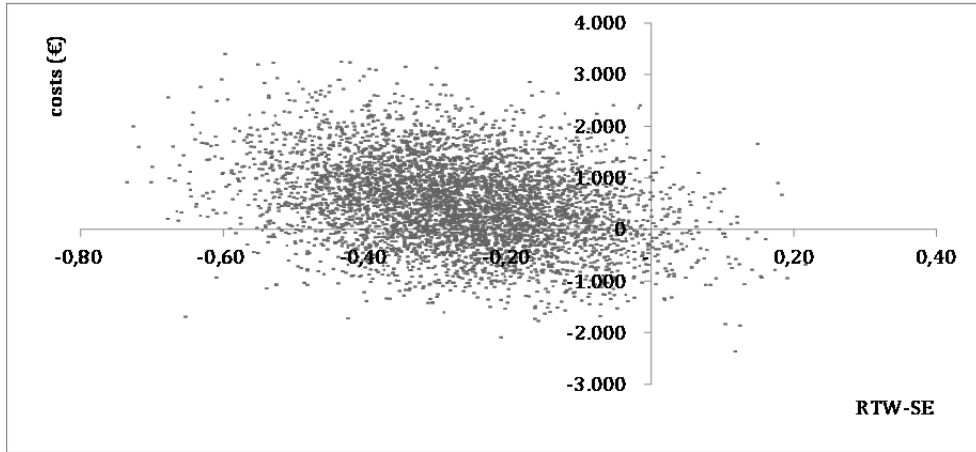
Table 2. Total costs per cost category and subtotal bootstrapped costs after 6 months follow-up

Cost category	Intervention group, costs €	Control group, costs €	Incremental costs, €	2.5-97.5 percentile
Intervention	0.48	0		
Health care utilization	405.84	318.58		
Medication and aids	113.36	104.92		
<i>Bootstrapped subtotal, mean</i>	<i>516.13</i>	<i>425.92</i>	<i>90</i>	<i>-195;475</i>
Patient and family				
Informal care	9.76	20.23		
Parking and travel	20.5	16.8		
<i>Bootstrapped subtotal, mean</i>	<i>30.03</i>	<i>36.63</i>	<i>-7</i>	<i>-28;12</i>
Other sectors (productivity)				
Absenteeism	2,577.43	2,176.82		
Presenteeism	57.36	38.68		
<i>Bootstrapped subtotal, mean</i>	<i>2,660.92</i>	<i>2,211.79</i>	<i>449</i>	<i>-943;1,894</i>
Bootstrapped total, mean	3,176.92	2,670.51	506	-1,076;2,063

Cost effectiveness

Table 3 shows that the differences in RTW-SE were minimal between the intervention and control group. A negative ICER of - 1,858.98 Euro was found, which means that the intervention yielded less effect though higher costs. The control condition dominates the intervention as 71% of the bootstrapped cost effectiveness pairs are located in the North West (fewer effects, higher costs) quadrant of the CEA plane (see Figure 2).

Figure 2. Cost effectiveness planes representing the uncertainty around the mean incremental costs and mean incremental effects of the intervention compared with the control condition.



Cost utility

The cost effectiveness plane from the CUA with a maximum WTP-level of €18,000 is shown in Figure 3. The probability of the intervention being cost effective was only 28% for a WTP threshold of €18,000. Overall, as figure shows, the intervention was less effective and more costly than the control condition (mean ICUR: -€354,726.08; see Table 3).

Figure 3. Cost effectiveness planes representing the uncertainty around the mean incremental costs and mean incremental effects of the intervention compared with the control condition and distribution of the cost-effect pairs above and below a maximum willingness-to-pay (WTP) level of €18,000.

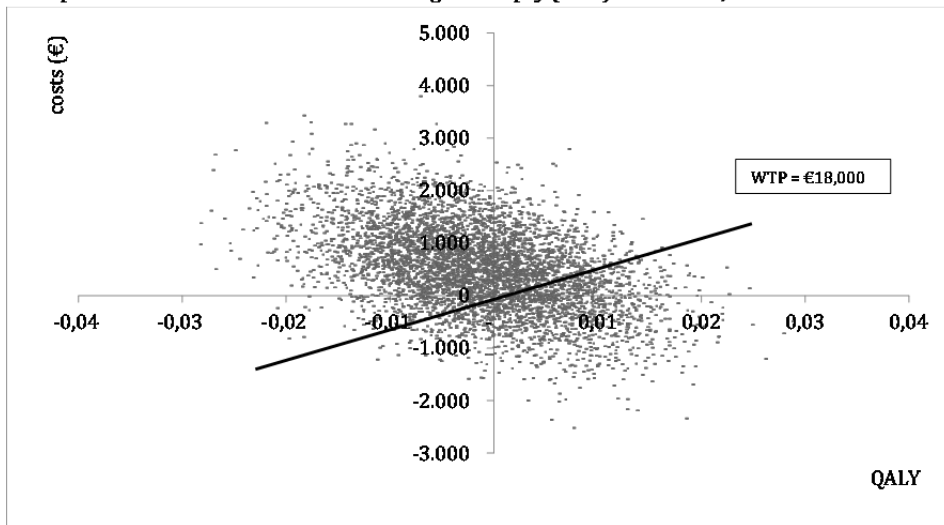


Table 3. Mean cost and effect differences between the intervention and control group including incremental cost effectiveness and cost utility ratios, and cost effectiveness plane distributions

Analysis	Effect measure	Sample Size		ΔCost Euro	ΔEffect	ICER/ICUR	Distribution (%)			CEA plane (quadrant)
		INT	CTR				NE*	SE‡ (dominant)	SW† (inferior)	
Cost Effectiveness	RTW-SE	113	128	508.64	-0.27	-1,858.98	1	2	26	71
Cost Utility	QALY	113	128	508.64	-0.0014	-354,726.08	25	18	9	48
Subgroup & Sensitivity										
Female GPs	RTW-SE	43	47	1,123.73	-0.3	-3,715.3	8	4	12	77
Male GPs	RTW-SE	70	81	169.94	-0.25	-674.17	2	5	39	54
Female patients	RTW-SE	70	72	89.52	-0.22	-412.25	4	6	38	53
Male patients	RTW-SE	43	56	1,393.09	-0.38	-3,703.75	3	2	17	78
Permanent employment	RTW-SE	91	102	-507.86	-0.19	2,611.86	1	8	65	26
Baseline differences	RTW-SE	113	128	-212.23	-0.15	1,385.77	6	13	47	34

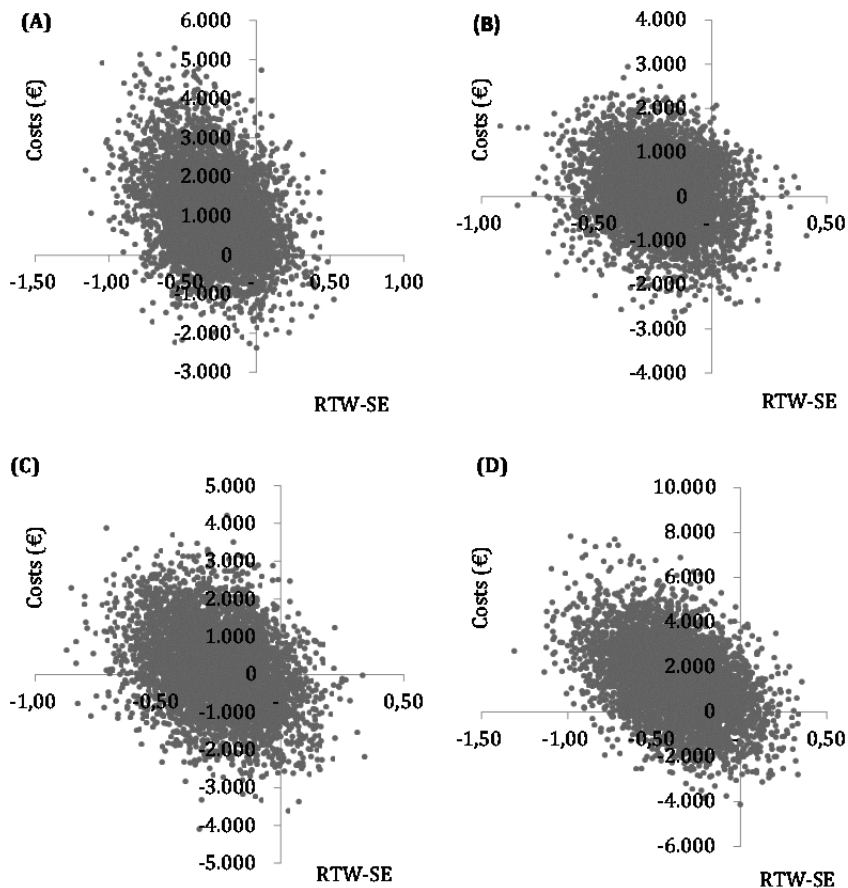
INT = intervention group CTR = control group

*NE refers to north-east quadrant of the cost effectiveness plane, which indicates the intervention is more effective and more costly than the control condition
‡SE refers to the south-east quadrant of the cost effectiveness plane, which indicates the intervention is more effective and less costly than the control condition
†SW refers to south-west quadrant of the cost effectiveness plane, which indicates the intervention is less effective and less costly than the control condition
‡NW refers to north-west quadrant of the cost effectiveness plane, which indicates the intervention is less effective and more costly than the control condition

Sensitivity analyses

Results of the sensitivity analyses are shown in Table 3 (bottom part). The cost effectiveness planes (CE-plane) representing the uncertainty around the mean incremental costs and effects for the four gender subgroups are shown in Figure 4. The first CE-plane showed that the intervention was not considered cost effective when studying RTW-SE among female patients and patients at consultation with female GPs. Regarding female GPs, an ICER of -3,715.3 Euro was found and most (77%) ICERs were located at the inferior quadrant. When only female patients are taken into consideration, differences in RTW-SE were small and the total costs were higher in the intervention group, indicating an ICER of -412.25 Euro. Most (about 90%) of the bootstrapped ICERs are also located at the inferior quadrant, indicating fewer effects. Gender subgroup analyses among male patients and male GPs supports the findings of the main analysis. Male GPs showed slightly more than half the cost effect pairs (54%) located in the inferior quadrant, indicating less effects (-0.25) for higher costs (€169.94) when comparing the intervention with the control condition. For male GPs, the majority of the ICERs (78%) are located in the North West (inferior) quadrant (large negative ICER).

Figure 4. Cost effectiveness planes representing the uncertainty around the mean incremental costs and mean incremental effects of the intervention compared with the control condition for the subgroups (A) Female GPs, (B) Female patients, (C) Male GPs, and (D) Male patients

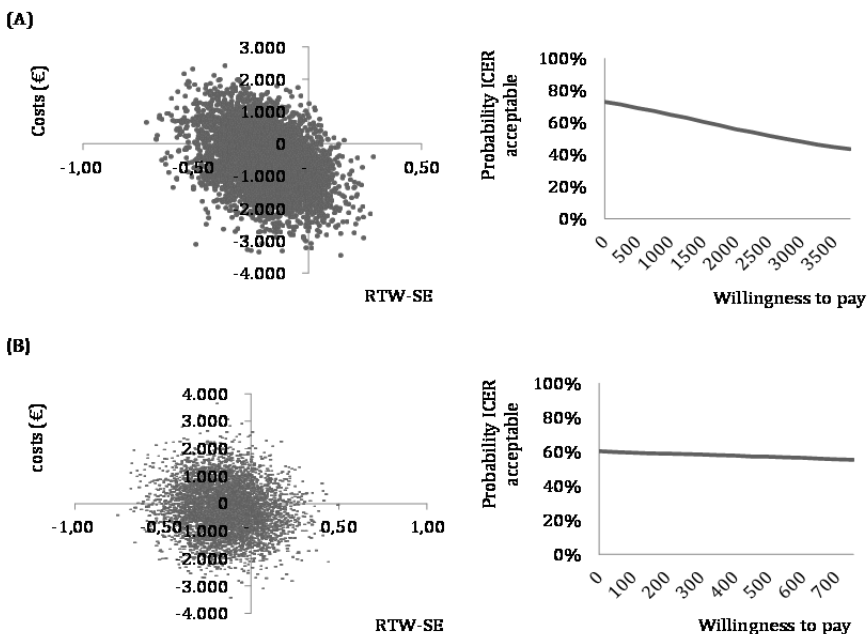


The cost effectiveness analysis for the subgroup ‘employees with permanent employment’ deviated from the main- and other subgroup analyses. For the subgroup ‘employees with permanent employment contracts’ (significant baseline difference), the intervention group yielded fewer effects (-0.19) on RTW-SE, albeit fewer costs (-€507.86) compared with the control condition, resulting in a positive incremental cost effectiveness ratio of almost €2,612. Sacrificing one point on the RTW-SE scale resulted in a cost saving of almost €508. Additionally, the CE-plane (Figure 5a) showed that 65% of the bootstrap cost effect pairs were located in the south-west quadrant, which indicates the intervention is less effective and less costly than the control condition. The cost effectiveness acceptability curve (Figure 5a) showed a 0.71 probability of the intervention being cost effective compared to

the control condition if one is willing to pay €250 for a one point increase on the RTW-SE scale. However, when increasing the WTP ceiling ratio, the probability of being cost effective decreases.

In the final sensitivity analysis, calculating the individual differences between the measurements, fewer effects (-0.15) on RTW-SE and lower costs (-€212.23) were found. As can be seen in Figure 5b, almost 50% of the ICERs were located in the south west quadrant (less effect, fewer costs). Figure 5b shows a slightly decreasing cost effectiveness acceptability curve and about 60% probability of the intervention being cost effective at a ceiling ratio of about €100.

Figure 5. Cost effectiveness plane and acceptability curves for the difference in RTW-SE after 6 months for (A) the subgroup employees with permanent employment and (B) adjusted for baseline differences.



Discussion

Main findings

The results of this study show that the training program aimed at increasing GP awareness to work-related problems, did not lead to improved RTW self-efficacy among patients in the intervention group when compared to the control condition. Furthermore, the cost difference was in favor of the control condition. In the incremental cost effectiveness plane, the majority of the ICERs fell in the quadrant of 'higher costs and less effects'. The gender-specific sensitivity analyses all confirmed

these findings. Based on the sensitivity analyses of the subgroup 'employees with a permanent employment contract' and the one accounting for baseline differences at the start of the study, the intervention resulted in less effects though fewer costs when compared to the control condition. In these analyses, the cost effectiveness plane the majority of the ICERs were located in the southwest quadrant (fewer costs, but less effective). Exploring the incremental cost utility resulted in findings comparable to the main cost effectiveness analysis (higher costs and fewer effects). Quality of life improved in both groups, though the difference between QALYs in both groups at six months follow-up was negligible (0.3). Overall, the lack of effectiveness might be due to the baseline values already being at a certain ceiling (RTW-SE scores around 5 and utilities around 0.8), making it difficult to improve.

Strengths and limitations

One of the strengths of this study is its generalizability as the study was conducted in routine daily clinical practice. Thereby, a broad generic population could be recruited which adequately represents patients visiting GP practices in the Netherlands and perhaps other countries with similar primary care systems. Furthermore, based on the characteristics of the intervention – and control groups, a successful randomization was achieved. Except for job contracts no significant differences between these groups were detected. Another strength of the study is the inclusion of costs related to both absenteeism and presenteeism in measuring productivity costs. Recent studies suggest that both measures are crucial since presenteeism attributes largely to the total productivity costs [29, 30]. Furthermore, the economic evaluation was carried out from a societal perspective. This enhances the possibility of using the results to allocate resources at a Dutch societal level and potentially to disaggregate data for future analyses from different perspectives (e.g. healthcare, employer, etc.). Furthermore, as previous research demonstrated the impact of gender on communication skills [11, 13, 14], this study provides in-depth insight into the cost effect pairs of the intervention group and the control group by including sub analyses regarding gender differences among both patients and GPs.

A number of limitations need to be considered as well. Firstly, it is possible that relatively motivated GPs were attracted by the training invitation. The fact that GPs were not blinded for the intervention (it was obvious whether they did or did not received the training) might have influenced their treatment approach underestimating the effect. Secondly, the cost and effect data were obtained via patient self-reported retrospective questionnaires which might have caused recall bias (the relevant period being six months). Recall bias could over- or underestimate true healthcare utilization and as such healthcare costs. However, as no 'gold standard' for measuring healthcare utilization is available, the method used in this study provides at least crude estimates of actual usage [31]. Due to practical and logistical reasons, the time span between both measurement points occasionally reached

over six months. However, the impact of potential recall bias was assumed to be equal for both groups due to the decent randomization procedure.

Concluding remarks

The intervention training program to increase GP awareness to work-related problems was not cost effective from a societal perspective with respect to both RTW-SE and QALYs outcomes. Based on these results the intervention could therefore not be recommended for implementation on cost-related grounds.

Acknowledgement

This work was financially supported by Foundation Institute Gak. We want to acknowledge Francine van den Driessen-Mareeuw for her indispensable assistance in collecting and cleaning the data. We also want to thank Hugo de Waal for his help in ridding our English of any Dutch and Belgian influences.

References

1. Weevers HJ, van der Beek AJ, Anema JR, van der Wal G, van Mechelen W. Work-related disease in general practice: a systematic review. *Fam Pract.* 2005 Apr;22(2):197-204.
2. van Dijk P, Hogervorst W, Buijs P, van Dijk F. Information in a Dutch GP registration system about high risks on long-lasting sickness absence: a cross sectional exploration in their registration system. *Eur J Gen Pract.* 2006;12(2):74-76.
3. OECD. *Sickness, Disability and Work: Breaking the Barriers: A Synthesis of Findings across OECD Countries.* 2010.
4. Wynne-Jones G, Mallen CD, Welsh V, Dunn KM. Rates of sickness certification in European primary care: a systematic review. *Eur J Gen Pract.* 2008;14(3-4):99-108.
5. Wynne-Jones G, Mallen CD, Main CJ, Dunn KM. What do GPs feel about sickness certification? A systematic search and narrative review. *Scand J Prim Health Care.* 2010 Jun;28(2):67-75.
6. Halvorsen PA, Edwards A, Aaraas IJ, Aasland OG, Kristiansen IS. What professional activities do general practitioners find most meaningful? Cross sectional survey of Norwegian general practitioners. *BMC Fam Pract.* 2013;14:41.
7. Waddell G, Burton AK. Is work good for your health and well-being? . In: TSO, editor. London2006.
8. Bakker R. *De samenwerking tussen huisarts en bedrijfsarts.* Groningen: University of Groningen; 2005.
9. Gezondheidsraad. *Beoordelen, behandelen, begeleiden: "Medisch handelen bij ziekteverzuim en arbeidsongeschiktheid".* Den Haag2005.
10. NHG. *NHG-Standpunt Kernwaarden huisartsgeneeskunde. Generalistisch, persoonsgericht en continu.* Utrecht: 2011.
11. Zaharias G, Piterman L, Liddell M. Doctors and patients: gender interaction in the consultation. *Academic medicine : journal of the Association of American Medical Colleges.* 2004 Feb;79(2):148-155.
12. Gulbrandsen P, Fugelli P, Hjortdahl P. Psychosocial problems presented by patients with somatic reasons for encounter: tip of the iceberg? *Fam Pract.* 1998 Feb;15(1):1-8.
13. ALM L-J. Medicine is not gender-neutral: influence of physician sex on medical care. *Ned Tijdschr Geneesk.* 2008 (152):141-145.
14. Sandhu H, Adams A, Singleton L, Clark-Carter D, Kidd J. The impact of gender dyads on doctor-patient communication: a systematic review. *Patient Educ Couns.* 2009 Sep;76(3):348-355.
15. de Kock KA, Steenbeek R, Buijs PC, et al. An education programme to increase general practitioners' awareness of their patients' employment: design of a cluster randomised controlled trial. *BMC Fam Pract.* 2014;15:28.
16. de Kock KA, Lucassen PJ, Spinnewijn L, et al. How do general practitioners perceive work-related issues of their patients? A focus group study. 2014.
17. Lagerveld SE, Blonk RWB, Brenninkmeijer V, Schaufeli WB. Return to work among employees with mental health problems: Development and validation of a self-efficacy questionnaire. *Work & Stress.* 2010;24(4):359-375.
18. EuroQol. EQ-5D-5L Value Sets [cited 2014]. Available from: <http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-value-sets.html>.
19. Matthews JN, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. *BMJ.* 1990 Jan 27;300(6719):230-235.
20. Statistics Netherlands. *Centraal Bureau voor de Statistiek. Voorburg/Heerlen*2012.
21. Hakkaart L, Tan SS, Bouwmans, CAM. *Manual for cost research. Methods and standards costs for economic evaluations in healthcare [in Dutch: Handleiding voor kostenonderzoek. Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg].* Rotterdam, the Netherlands: Instituut voor Medical Technology Assessment, Erasmus Universiteit Rotterdam, 2010.
22. Healthcare institute Netherlands. *Pharmacotheapeutic compass [in Dutch: College van Zorgverzekeringen. Farmacotheurapeutisch kompas].* [Cited September 2012]; Available from: <http://www.fk.cvz.nl/>.

CHAPTER 10

23. Zorginstituut Nederland. Drug costs in the Netherlands [cited 2014]. Available from: <http://www.medicijnkosten.nl/>.
24. Zorginstituut Nederland. 2014 [cited 2014]. Available from: www.gipdatabank.nl.
25. Koopmanschap MA. PRODISQ: a modular questionnaire on productivity and disease for economic evaluation studies. *Expert Rev Pharmacoecon Outcomes Res.* 2005 Feb;5(1):23-28.
26. Koopmanschap MA, van Ineveld BM. Towards a new approach for estimating indirect costs of disease. *Soc Sci Med.* 1992 May;34(9):1005-1010.
27. Statistics Netherlands. Jobs (outstanding, new and filled) [cited 2014]. Available from: <http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=80857NED&LA=NL>.
28. Van Asselt AD, van Mastrigt GA, dirksen CD, et al. How to deal with cost differences at baseline. *Pharmacoeconomics.* 2009;27(6):519-528.
29. Noben C, Smit F, Nieuwenhuijsen K, Ketelaar S, Gartner F, Boon B, et al. Comparative cost-effectiveness of two interventions to promote work functioning by targeting mental health complaints among nurses: pragmatic cluster randomised trial. *Int J Nurs Stud.* 2014 Oct;51(10):1321-31.
30. Krol M, Papenburg J, Koopmanschap M, Brouwer W. Do productivity costs matter?: the impact of including productivity costs on the incremental costs of interventions targeted at depressive disorders. *Pharmacoeconomics.* 2011 Jul;29(7):601-619.
31. Noben C, de Rijk A, Nijhuis F, Kottner J, Evers S. Patient self-reported and administrative data used for cost estimations of healthcare resource use: A systematic review of methodological reporting quality and validation evidence. 2014.

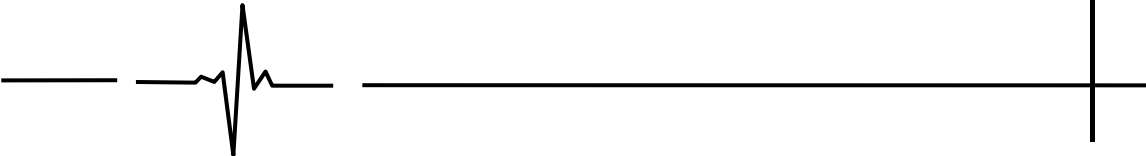
Appendix 1. Intervention costs

Type of cost factor	Calculation	Total Cost (€)
Personnel costs		
Group training expert speaker (N=2)	<i>Time investment (10 hours)*hourly wage (€100)</i>	1000
Booster training expert speaker (N=2)	<i>Time investment (6 hours)*hourly wage (€100)</i>	600
Preparation and travel time (for 2 trainers)	<i>Time investment (16 hours)*hourly wage (€100)</i>	1600
Total personnel costs		3200
Material costs		
Binders and materials	<i>Binders & materials (€15)*number of attendees and trainers (14)</i>	210
Training location (incl. meals and beverages)	<i>(costs location group training €1317,75) + (costs booster training €122,15/2)</i>	479,97
<i>Total material costs</i>		<i>689,97</i>
Total intervention costs		3890
Per Patient Intervention Costs	<i>total intervention costs / patients being served by the intervention (N=8074)*</i>	0,48

*average number of working patients per GP practice (672,8) multiplied with the number of GP practices included in the intervention (12) (per trainer approximately 6 GPs in one group, per training 12 GPs)

Part IV

DISCUSSION



CHAPTER

11

General discussion

Discussion

Information on the cost effectiveness of employability interventions for employees with health problems guides towards priority-setting in the field of occupational health. Methodologically sound economic evaluations play an important role in decisions about the development and implementation of these employability interventions. The starting point for this thesis was that further development and refinement of both employability interventions and the methodologies for economic evaluation in the field were needed. The objective of this PhD-thesis was to provide insight into both methodological and empirical (experimental) analyses of interventions aimed at achieving sustainable employability for employees with health problems. The thesis therefore provides insight into (i) the methodological challenges, and (ii) the application of this methodological knowledge to different employability interventions to generate economic evaluation results.

In Section I of the thesis, measurements for health-related productivity and resource use were critically assessed to improve the methodologies for economic evaluation. Furthermore, transparent information on potential methods to conduct economic evaluations in the field of occupational health was provided (**Chapter 2-4**). In the Section II and Section III, this methodological knowledge has been applied to several existing (**Chapters 5-7**) and innovative (**Chapters 8-10**) interventions aimed at achieving sustainable employability for employees with health problems. The methodological knowledge has been applied to generate new evidence on their cost effectiveness.

This final chapter presents and discusses the main findings of the studies. Furthermore, it reflects on practical and methodological considerations, and provides implications for future research and practice based on the experiences and conclusions.

Main findings

(1) The methodological challenges when measuring health-related productivity and resource use

A useful starting point to examine the state of affairs within a research field is to conduct a systematic review. A systematic review is a literature review focusing on a research question aimed at identifying, appraising, selecting, and synthesizing the research evidence relevant to the research question [1]. For this PhD-thesis, two systematic reviews have been conducted.

In the first systematic literature review (**Chapter 2**), the measurement properties in generic self-reported instruments measuring health-related productivity changes were critically appraised and compared [2]. Twenty-five articles were included, reporting on the reliability, validity, and responsiveness of 15 different instruments. The methodological quality appraisal was assessed using the COSMIN-

checklist [3, 4]. The analyses showed that information on cross-cultural validity and information on measurement error was often lacking. Aside from evidence on the psychometric properties of the Work Limitation Questionnaire (which was the most frequently assessed instrument), the Stanford Presenteeism Scale, and the Productivity and Disease Questionnaire, the overall evidence of the other 12 instruments was limited and of fair to poor quality. The review concluded that high-quality studies providing details on the measurement properties of instruments measuring health-related productivity were lacking. In addition, transparent information to aid deciding on the appropriateness of the instrument was frequently not reported (e.g. information on the population of interest, the content of the instrument, the purpose of its use, and the target country).

In the second systematic literature review (**Chapter 3**) the validity of self-reported and administrative healthcare utilization tools for cost estimation in economic evaluations was assessed. A quality appraisal tool, derived from existing guidelines, was used to evaluate the methodological reporting quality of studies comparing the validation evidence of healthcare resource use measurement tools. Out of the 16 included studies, only six studies scored 'adequate' on the minimal criteria for reporting, and could be considered trustworthy. However, no evidence was established that one measurement tool (measuring self-reported versus administrative healthcare utilization) is better than the other. Next to both random and systematic errors, varying levels of agreement among the different data sources were found in the included studies. The validation evidence of the small number of studies with adequate methodological reporting quality cautiously supported the exchangeability of both the self-reported and administrative resource use measurement methods.

The study protocol (**Chapter 4**) provided information on possible methods used to conduct an economic evaluation (i.e. cost effectiveness, cost benefit, and cost utility) of employability interventions aimed at achieving sustainable employability for employees with health problems [5]. The comparisons between the employability interventions and the usual care- or control conditions were pragmatic in nature; replicating the actual ((occupational) health care) setting in which the studies occur. The design is compliant for different settings, for example when different unit prices for resource utilization, or divergent outcome measures, are preferred.

Given the results of the studies included in Part I of this PhD-thesis, the economic evaluation of existing- and new employability interventions are based on solid methodologies, accounting for stakeholder interests, and proper methodological reporting quality to ensure transparency. The economic evaluation results are summarized in the paragraphs below.

(II) The economic evaluation results of existing interventions aimed at achieving sustainable employability for employees with health problems

In the second part of this PhD-thesis, the economic evaluation results of two different employability interventions are presented.

In **Chapter 5**, an economic evaluation alongside a pragmatic cluster randomised controlled trial was conducted to evaluate the comparative cost effectiveness of two strategies to promote work functioning among nurses by reducing symptoms of mental health complaints [6]. Three conditions were compared in two sub-economic evaluations. After six-months follow-up, the comparison of the occupational physician condition group (existing of screening, feedback and referral to the occupational physician for screen-positive nurses) with the control condition group (consisting of online screening for mental health problems without feedback) resulted in a dominant incremental cost effectiveness ratio. The comparison yielded a cost saving of €5,049 per treatment responder and could therefore be recommended for use in practice. Treatment response was defined as an improvement of at least 40% on nurses' work functioning between baseline and after six-months. The second comparison was made between the control condition group (screening without feedback), and the e-mental health condition group (whereby screen-positive nurses were provided with feedback regarding their screening results and referred to an e-mental health intervention). This resulted in less treatment response and added costs per treatment responder (about €4,000) for the e-mental health condition. Because the e-mental health condition was associated with a higher probability of fewer benefits at higher costs, compared to the control condition, this employability intervention was recommended to not be put in practice.

A more in-depth analysis was conducted for nurses at elevated risk of mental health complaints and lost productivity (**Chapter 6**). To evaluate the balance between the costs of a preventive intervention and the cost offsets stemming from improved productivity from a business perspective, a cost benefit analysis was studied. Within six months, the intervention costs of personalized feedback and occupational physician utilization were more than recouped, representing a favourable business case from the employer's perspective. A return-on-investment of up to €11 per invested Euro per nurse was found.

A pilot cost effectiveness study evaluated how economic evaluation results contributed to the development and implementation of a web-based employability intervention (**Chapter 7**). The pilot cost effectiveness study assessed a web-based employability intervention aimed at increasing self-perceived work ability among employees who experienced problems in relation to their employability. This pilot economic evaluation demonstrated the value of cost effectiveness analyses preceding a full-scale study. Both the content and the study-design of the web-based employability intervention were critically considered. Content related improvements were required. Firstly, low uptake and compliance rates, particularly for the online intervention element (3%), were found for the web-based employability interven-

tion. Secondly, additional analyses concluded that the intervention appeared to general for specific situations, such as prolonged health problems requiring specific questions on work and health. Thirdly, the originated tendency that females and participants aged 50 or older are more sensitive for an intervention effect supported the need for better targeting of the web-based intervention. In general, the web-based employability intervention was not associated with an economically relevant impact over regular trade union support for a heterogeneous population with work- and disability related concerns. Based on the results of the pilot economic evaluation of the web-based employability intervention, also the study design required adaptations. The mean incremental differences on work ability and quality adjusted life years were unable to demonstrate an effect of the web-based intervention (respectively -0.51 and -0.01 point improvement). It is therefore disputed whether the most appropriate outcome was selected to assess the true effect of the intervention, or whether the effect of the intervention was influenced by other factors (for instance derived information other than via the web-based employability intervention). Well-thought outcome measures are needed in future studies for this context. In addition, based on the relative contribution of diverse cost categories with respect to the total overall costs, cost category measurements require revision as well. Based on the negligible relative contribution of patient- and family out-of-pocket costs to the total overall costs, data gathering for these cost categories could be omitted to minimize the burden for the employees and their environment.

(III) The economic evaluation results of innovative interventions aimed at achieving sustainable employability for employees with health problems

In the third part of the thesis, the economic evaluation results of three newly developed employability interventions are presented.

The cost effectiveness, cost utility and cost benefit of a new organizational return-to-work intervention to improve Cooperation between Sick-listed employees and their Supervisors (COSS) was carried out alongside a field study with a six-month follow-up (**Chapter 8**). The main outcomes were the number of days until return to work, quality adjusted life years, and productivity gains. The number of days until return to work was conceptualized as the first moment at which progress is made in hours worked, even if the employee works less hours than contracted or if the employee is still (partly) sick listed. When compared to common practice, employees in the COSS-group reduced almost 16 calendar days of absenteeism and saved approximately €27 per day of work resumption. The cost utility analysis also exhibited reduced costs, however the gain in quality adjusted life years was uncertain. From an organizational perspective, the net benefit of COSS compared to common practice yielded a productivity gain of almost €396 per invested Euro. Implementing COSS for sick-listed employees has a large potential to reduce societal and organizational costs and to improve productivity among sick-listed employees.

A short-term economic evaluation was conducted to provide first insights into the cost effectiveness of an intervention program aimed at improving work productivity for workers with rheumatoid arthritis (**Chapter 9**). The intervention coordinated clinical- and occupational care at the worksite to support work participation for workers with rheumatoid arthritis. Although not significant, at-work productivity loss was about 0.5 hours higher in the intervention group, and the mean QALY was on average 0.01 point lower when compared to the control group. Neither one was preferable in terms of costs. After six-months follow-up the costs in the intervention group were approximately €500 higher compared to the control group. Based on the six-month economic evaluation results, no evidence was found to support the implementation of the intervention.

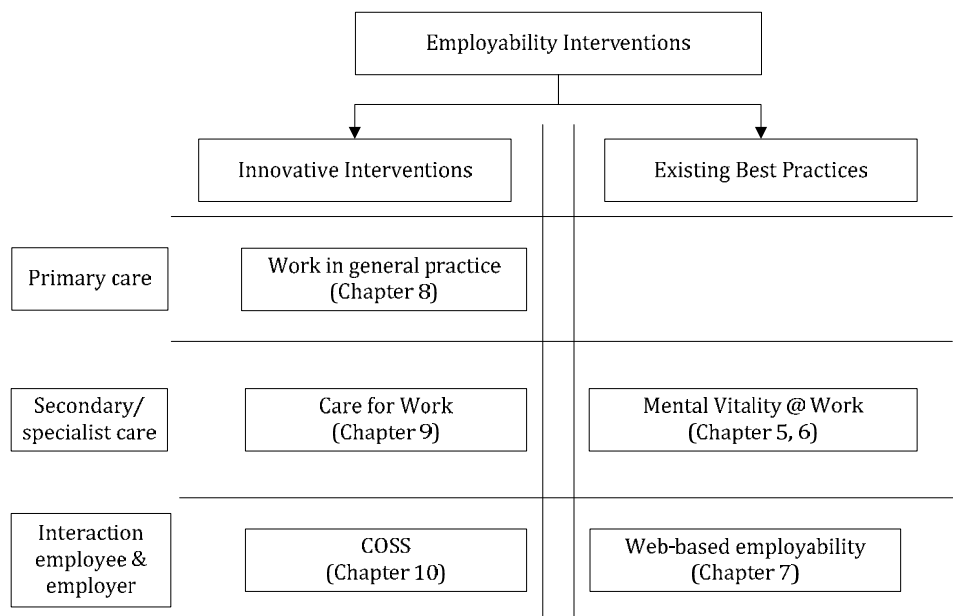
A final economic evaluation was conducted for a tailored educational training program for general practitioners with the aim of increasing their awareness of work-related health problems (**Chapter 10**). This training program aimed to improve treatment strategies in the primary care setting by increasing work-related self-efficacy (measured with the return-to-work self-efficacy scale) and quality of life for employees. Differences in the main outcome measures between the intervention group and the usual GP-care group were minimal. A negative incremental cost effectiveness ratio revealed fewer effects at higher costs for both the cost effectiveness and –utility analyses. Based on the six-month economic evaluation results, no evidence was found to support the implementation of the intervention.

Main considerations

Practical considerations

At the start of the study several practical difficulties occurred when recruiting practices to participate. Although nine practices initially showed interest in participating in the study, few of them were either unable to provide the minimum required number of participants; did not want to have their intervention assessed in a comparative design; or were unable to provide a comparable control group. At the start of the PhD-study, five employability interventions were planned to be assessed (see Figure 1, Chapter 4)[5]. Throughout the trajectory, several practical difficulties recurred, slowing recruitment of eligible participants, and consequently the data collection. The initial selected existing practices involved in the PhD-study were unable to recruit the required number of participants. Therefore, a number of other initiatives had to be instigated. Figure 1 describes the included employability interventions and the corresponding chapters wherein the economic evaluation results are presented in this PhD-thesis.

Figure 1. Employability interventions whereof the economic evaluation results are included in this PhD-thesis



Another practical consideration when conducting the economic evaluation of employability interventions was the follow-up period. Ideally, the follow-up time was 18 months because after this point in time the results are expected to extend into the subsequent years [5]. However, because of the recruitment difficulties during the study period, an 18-month follow-up period was not achieved. The economic evaluation results presented in this PhD-thesis therefore compare the employability interventions (in terms of costs, effects, and utilities) to usual care over a timespan of six months. The ‘Mental Vitality @ Work’ (Chapters 5 and 6) and the ‘COSS’ (Chapter 8) studies are completed and the economic evaluation results are published or submitted for publication. When preparing and discussing possibilities to conduct the economic evaluation of the web-based employability intervention (Chapter 7), the minimum required number of participants could not be reached. Because both the researchers and the practices wanted to learn from the research, a different research approach was followed - a pilot economic evaluation. For the remaining two interventions (‘Work in General Practice’ and ‘Care for Work’), the data collection is still ongoing, and consequently, the economic evaluation results included in this dissertation thesis are interim-analyses (**Chapters 9 and 10**). The possible directions for decision-making will be explored below, even though a short follow-up time had been used for the analyses.

Methodological considerations

According to the health economic guidelines it is recommended to explicitly state the chosen perspective in each study as it influences the costs and effects to be measured [7, 8]. Furthermore, economic evaluations are strongly recommended to focus on the society as a whole, which is translated in terms of 'the societal perspective'. By presenting all socially relevant costs and effects in a disaggregated form, the findings can be generalized to other settings [7, 8]. Subsequently, the analyses can be narrowed by solely focusing on those costs and effects that affect the interested stakeholders, thus, conducting the analysis from another perspective. For example, when conducting the economic evaluation of a health care intervention, healthcare providers are the interested stakeholders. From their perspective and only those costs related to healthcare utilization would seem relevant. It is clear that when conducting the economic evaluation of a healthcare intervention from the provider's perspective, by solely measuring and valuing healthcare utilization costs, the broader societal perspective cannot be applied afterwards. The societal perspective in fact also assumes the incorporation of productivity costs, patient out-of-pocket costs, and those costs that would be difficult to incorporate at the end of the study. Throughout this PhD-study resource use is self-reported by participants and collected separately from the unit costs or prices of those resources. Collecting resource utilization separately provides opportunities to apply health care- and intervention- provider specific prices to the units of resource use.

In the field of occupational health, the perspective of economic evaluative studies is just as essential as in general health technology assessments. However, in order to be as complete as possible, this field requires a more narrow view than the societal perspective, and a broader one than the focus of the organization or employer. Different perspectives take different costs and consequences into account, depending on the stakeholder's interest. Therefore, explicitly reasoning and flexibility in the choice of study perspective is warranted. To illustrate this, investing in employability interventions whereby the benefits (for example quality of life) are entirely for the employees, without impacting productivity, can be justified from a healthcare perspective. Investing in this 'QALY increase' can also be justified in social terms, but might not seem that relevant from the employer's perspective who is mainly interested in productivity changes. Although there is a strong case for focusing on a wide range of health-related outcomes (such as QALYs), financial outcomes are also important. Whether human or financial outcomes are more important is another discussion, and not addressed in this PhD-thesis. The issue here is that the perspective of the study should be determined in collaboration with the relevant stakeholders, and if needed, be supplemented with analyses incorporating additional or divergent effects and/or costs (such as a return-on-investment analysis from the organizational perspective).

Elaborating on the study perspective also stresses the inclusion of relevant cost categories in this field of study. The answer to the question which costs to include always depend on the chosen study perspective [9]. Including all costs despite of

who bears or benefits them, runs the risk of double counting. The consequences of employability interventions are often expressed in terms of changes in health-related productivity [10, 11]. Including all costs also implies the inclusion of changes of health-related productivity into a financial metric [12]. It is clear that when health-related productivity has to be included as a cost parameter in order to comply with the guidelines [7, 8], they cannot serve as the outcome measure. Adjusting the preferred societal perspective (according to the health economics) as presented in the guidelines [7, 8], is an important methodological consideration that should be taken into account [12]. Where possible throughout the studies included in this PhD-thesis, double counting was avoided in the cost effectiveness analyses by calculating and summing intervention-, health care-, and patient- and family- costs in the cost effectiveness analysis. When productivity served as the outcome of interest (i.e. at-work productivity loss), the costs of changed productivity were omitted. Hence, a full societal perspective could not be applied.

When assessing interventions aimed at achieving sustainable employability among employees with health problems, the quality of the measurement method and the valuation method to quantify costs influence the usefulness of the economic evaluation results [2, 13]. The debate continues on what is the most reliable and best method for collecting costs in economic evaluations [14, 15]. Because there is no consensus, data collection is generally guided by pragmatic concerns, such as the resources available or the characteristics of the populations under research. Nonetheless, more attention needs to be paid to both measuring and valuing costs. The measurement tools used for cost-data collection (health-related productivity and healthcare resource use) would benefit from high-quality methodological reporting in order to trust the validation evidence forehanded. Practical and applicable knowledge is best to be considered with insight into the reliability, validity and responsiveness of instruments used to evaluate cost-data [2, 3]. This PhD-thesis found that the methods used to measure resource utilization costs resulted in inconclusive results. The validation evidence of a (small) number of studies with adequate methodological reporting quality only cautiously support the exchangeability of both self-reported and administrative resource use measurements. The reason for this is essentially the paucity of coherent methodological quality hampering the interpretation of estimates and collected validation evidence. Because the provision of healthcare services has become more integrated, economic evaluations need to include information from many different healthcare providers. However, healthcare utilization is not always registered in social security systems (e.g. alternative care), hampering the calculation of economic costs. Therefore, a robust, uniform and transparent method is required.

This thesis also found that there is a need for adequate measurement and valuation of productivity costs. Currently, sick leave days are most commonly used [16]. However, no certainty can be provided to whether this proxy accurately reflects the true cost related to lost production (as not all sick leave days result in lost productivity). In all economic evaluations included in this PhD-thesis, costs stemming from

changed productivity incorporated both the number of days absent from work (absenteeism) and the number of workdays lost due to decreased work performance (presenteeism). To value productivity losses, three principals could be used. One is based on the human capital approach. To value the costs, the complete lost production time is incorporated [17]. The second is based on the friction cost approach, which states that productivity loss only occurs until the tasks of the absent employee are compensated for. After a fixed friction period, it is assumed productivity has restored to its maximum and no more costs are attributed to the absent worker [17]. The third method assumes that productivity costs only arise when lost work is not compensated for, for example by colleagues or through extra hours [18]. The three methods risk the production of widely different results. The human capital approach assumes continuous productivity costs. The friction cost approach applies only during a fixed period until the production is restored. The third approach assumes that changed productivity compensated for does not result in any lost productivity costs. The latter approach is hard to put in practice in the Netherlands. Even compensated productivity losses are not costless since every employer in the Netherlands has the obligation to continue paying salary for two years when employees are absent from work due to health problems [19]. This law was introduced to stimulate the employers' responsibilities for reintegration of sick employees (Gatekeeper Act). Dutch employers must continue to pay at least 70% of the salary of sick employees for the first two years⁴. Within this period, productivity might be restored in terms of output, when, for example, the sick employee was replaced by an outplacement worker. The costs for the employer increase then because an 'extra worker' has to be hired to restore productivity, on top of the wage payment obligation the employer has towards his absent employee for the upcoming two years [19]. This insight is important since the perspective of the employer has become more prominent as a result of a number of legislative developments. The generalizability of the findings in The Netherlands towards other countries however might therefore be limited by legislative differences.

Applying the friction cost method to calculate productivity costs due to health constraints among employees in countries other than the Netherlands require adjustments for country-specific social security arrangements. As illustrated above, disability pensions are only rewarded after two years of sick-leave and sufficient reintegration efforts. In the Netherlands, the friction period was estimated to be about 13 weeks⁵. Generalizing this approach to another country, with a different

⁴ Most employers top up the wage payment from 70% (legally required) to 100% on the basis of collective agreements; so a supplement of 30% is customary in the first 52 weeks. In the second year, this supplement is no longer customary.

⁵ The following formula was used to calculate the friction period in weeks: $FP = \left(\frac{ov}{cv} * 12\right) * 4.35 + 4$. In the equation, *ov* represents the average number of open vacancies in one year and *cv* represents the average number of completed vacancies in one year. The average vacancy duration in months is converted to weeks with the 4.35 conversion factor. The added 4 weeks estimate the period which supervisors are

social security system where disability pensions are paid sooner and the duration of the friction periods vary, will most likely result in divergent outcomes.

Additionally, increased attention for the relative contribution of different cost categories with respect to the overall costs is preferred. This would ultimately provide insight into the relevance of measuring the costs from a societal perspective as recommended by the guidelines [7, 20]. Furthermore, there are other methodological considerations with respect to the outcome measures selected. As previously mentioned, the choice of outcome (for instance days of productivity versus QALYs) in this field of study is often influenced by the type of economic evaluation and the risk of double counting. Properly addressing a priori usage and identification of items for effect estimations is essential. Productivity consequences of health improvements can be measured as an effect of the intervention. However, it is important to also look at a broader range of effects available to employees and society as a whole. To identify the full extent of consequences, QALYs can be used. Including QALYs improves the generalizability of employability interventions by enabling potential comparison of economic evaluation results of various general health interventions [21]. When conducting a cost utility analysis whereby QALY serves as the outcome of interest, the costs in other sectors (productivity costs) can be included without risking double counting. Although several studies in this PhD-thesis used utility measures of health changes associated with employability interventions, the QALY was not always sensitive enough to detect an intervention effect. A QALY combines health gains from reduced morbidity and mortality into a single measure based on the 'desirability' of the different outcomes [21]. When using QALYs as a means to gather economic evaluation results for employability interventions among employees with health problems, a QALY must satisfy a constant proportional trade-off between quality and quantity of life, and must be independent of disease incidence and impact [21]. It can be argued that this is discriminating against work-disabled people or employees with health problems impeding their labor participation. Therefore, after considering this 'equity distribution', a sensible use of the technique and adequate interpretation of findings is needed. Even more so, the studies in this PhD-thesis which used QALYs as an outcome measure, showed high baseline utilities, inducing ceiling effects which makes it difficult to detect substantial improvements [22]. Nonetheless, insight into changes in the quality of life due to the employability intervention produces useful information to aid decision-making in the field. Additionally, it is recommended to include the effect of work disability on productivity in a quality of life instrument [23]. Respondents do not include the effect of health constraints on productivity if the measurement instrument used does not explicitly ask about this effect [23]. Potential reasons for the discrepancy between work disability and QALY scores should be

supposed to use before deciding on temporary or permanent replacement of the absent worker. Based on the formula, the friction period for 2012 in the Netherlands contained 13.24 weeks (92.68 days).

further assessed. This will attribute to gain insight into the psychometric properties of quality of life instruments (e.g. EQ-5D-5L) among employees with health problems, especially in a disease-specific population (such as for example patients with rheumatoid arthritis).

Considerations regarding the interventions

Several employability interventions in this PhD-thesis were unable to detect a substantial improvement in their main outcome measures. This however does not imply the employability interventions are not meaningful in general. Considerations regarding the interventions have to be taken into account though.

A first consideration might postulate that usual work resumption in the Netherlands is already of high quality; that is, the control condition was already likely effective, which limits the visibility of the improvements in outcome due to the intervention. Another consideration could be the poor implementation of the employability interventions (or components of the employability intervention).

In addition, some effect measurements do not represent what is truly relevant to expect of an employability intervention for employees with health problems. For example, generic QALYs are not sensitive enough to assess the effect of an employability intervention for employees with a diverse range of health problems that impede their labor participation [22]. Nonetheless, the added value of each employability intervention should be assessed properly, clearly stated and communicated to all stakeholders involved. Economic evaluation results are extremely valuable for informing judgements on the 'worthiness' of investing in employability interventions.

Although all interventions in this thesis were developed to improve employability amongst employees with health problems, the included studies induced that the field of occupational health is not that easily affected, even when the economic evaluation results of the employability interventions are relatively positive. Adequate implementation of the employability intervention in practice requires meeting the needs of all stakeholders. Stakeholders affected by the intervention have concerns about the distribution of costs and consequences, and have divergent priorities and incentives. To illustrate this, a broad range of costs and consequences across the different parties involved has to be considered to express the distributional implications of the alternatives. At first sight, the investment costs of an employability intervention are primarily allocated with the employer and the benefits of, for example health and quality of life do not appear relevant. However, the economic evaluation results might show contrasting and interesting benefits for the employees, at minor investment costs, making it worth to invest in the employability intervention. Hence, decision making requires the acknowledgement of context-specific characteristics, which make the employability intervention more or less relevant to particular stakeholders.

Aside from full economic evaluations, pilot economic evaluations are useful to investigate the reach, compliance and perceived cost effectiveness of an employa-

bility intervention before implementation occurs. Pilot economic evaluations re- sults attribute to the intervention development. Valuable resources can be saved when adjustments in content and design can be made before large-scale studies are set up [24-26], and this PhD-thesis shows that this also applies to economic evalua- tion studies. Insight into how and where to target the intervention, given its popula- tion of interest, would be very valuable.

Implications for practice and future research

The findings described throughout this PhD-thesis have implications for practice and future research as they contribute to the methodological and practical debate concerning the need for sustainable employability among employees with health problems.

When poor methodological reporting quality is at stake, there is a risk that the economic evaluation results are biased, which lead to inappropriate decision- making. To illustrate this, there is no common reference standard for measuring resource use. Furthermore, based on the findings of systematic review (Chapter 3), only six studies were found to support the exchangeability of self-reported and administrative healthcare resource use measurements. This could result in the risk of misleading conclusions. Furthermore, the methods for measuring and valuing health-related productivity costs induce the risk that results could be biased, de- spite a large number of measurement tools that have been developed over recent years to measure changes in work productivity due to health-related absenteeism and presenteeism [2, 10, 11, 27]. Both systematic reviews included in this PhD- thesis established that there was a lack of details on the measurement properties resulting in low methodological reporting quality. This impedes the decision mak- ing for choosing an appropriate instrument.

Although economic evaluations require assumptions to be made, more focus on testing these uncertainties in sensitivity analyses would help to present the robust- ness of the results under alternative scenarios. Conducting an economic evaluation of employability interventions in future research can justify the choice of candidate variables for the sensitivity analysis. Preferably those variables with an impact on the economic evaluation results and inconclusive measurement properties (such as changes in health-related productivity and resource utilization costs) should be included. In future research more attention could also be paid to subgroup analyses to account for participant heterogeneity, which provides valuable information on how to target the interventions aimed at achieving sustainable employability for employees with health problems.

In addition more focus on characterizing the uncertainty around the joint cost- effect estimates (CEA-plane) is preferred. At the moment the precision of the cost- effect estimates is quantified and non-parametric bootstrapping techniques deal with the highly skewed nature of cost data [28]. However, no visual methods (such as the cost effectiveness acceptability curve), have yet been developed to help oc-

cupational health decision makers interpret the results based on a uniform willingness-to-pay or willingness-to-accept for increased health-related productivity.

Furthermore, more insight into the relationship between health consequences and productivity consequences, that is, health-related productivity, is warranted. For example, health utility measures do not yet explicitly consider productivity in the elicitation questions asked when assessing QALYs in occupational health research [23]. Assigned utility weights on quality of life are debatable for different health conditions, and can thereby impact on productivity. Additionally, this also relates to the issue of double-counting, as it remains unclear whether productivity changes are to be addressed; as a cost- or outcome parameter, and how to avoid double counting. Since incremental cost effectiveness ratios are sensitive to what is placed in the numerator and denominator [29], a clear description of the content of both the numerator and denominator is needed to meaningfully compare the results of different studies.

Awareness into compensation mechanisms is also required. Changes in productivity due to health problems are compensated because colleagues take over the work, or workers themselves make up for the lost work after their return. Despite uncertainty regarding the valuation methods to be used, more research on factors influencing changed productivity and the measurement of productivity costs would be helpful to interpret the results. For example, the type of job is important, as the production process nowadays is more individually oriented. Future research could evaluate productivity changes across a variety of jobs and workplaces, requiring adequate methodological reporting quality with respect to the psychometric properties of the measurement tools used.

Next to that, clear insight into the intervention costs aids insights in the content of the intervention. Distinctions between human and material resources help to unravel the precise content of the intervention. Understanding the mechanisms of the intervention is helpful to better appreciate why parts of an intervention are poorly implemented or not adequately tailored to the target participants. To illustrate this for purpose of cost calculations, material resources are reported in capital items with a time span longer than one year (e.g. equipment needed to build the interventions), and recurrent or revenue items consumed in less than one year (e.g. maintenance costs). This information can positively affect content related improvements regarding the intervention. When minimal maintenance costs throughout the year are detected, this could induce a minimal of 'updates' regarding the intervention in order to keep it tailored. As an example, e-health requires renewals to correspond with continuous software updates. If maintenance costs are minimal, it can be assumed that no updates occurred over the last year, hampering adequate implementation of components of the interventions, and thereby impeding adequate targeting.

Additionally it is often assumed cost effectiveness analyses are only useful after complete intervention development, and not in the exploratory phase of intervention development. However, because economic evaluations add new perspectives,

cost effectiveness analyses are also very useful to contribute to the development of the intervention, and – as in the scientific literature regarding pilot studies is emphasized – the study design [24-26]. In all, information from economic evaluations can help inform decision-making. However, the economic evaluation results of the different existing and innovative interventions aimed at achieving sustainable employability for employees with health problems indicated difficulty adhering to all stakeholders' interests. The trade-off that matters most to the stakeholders affected by the intervention, determined the type of economic evaluation chosen. Throughout this PhD-thesis, mainly cost-benefit, cost-utility, and cost-effectiveness analyses were conducted. Cost benefit analyses were more relevant from the employers' perspective as it provided a direct assessment of the impact of the investment in employability on productivity changes for the organization. Cost utility-, and cost effectiveness analyses were more relevant to represent the interest of workers and healthcare providers or other stakeholders (e.g. health insurers) because of the related health outcomes. Nonetheless, the costs and consequences of employability interventions are distributed across multiple involved stakeholders, future insight therefore should include the reporting of both monetary- (productivity changes) and non-monetary (health, work-ability, QALY, etc.) outcomes for public and organizational consequences. Future research should allocate the reporting of consequences in multiple ways and separately. Additionally, an aggregate summary measure should be included, in order to highlight the intervention's impact on both types of consequences.

An innovative method to do so is by conducting a so-called 'social cost- benefit analysis' (*MKBA in Dutch*) [30]. The 'MKBA' is an information tool that systematically portrays and values the advantages and disadvantages of an intervention from the perspective of society as a whole and supports the policy choices regarding the implementation of an intervention, wherever possible, on objective grounds. The MKBA is applicable to all types of interventions in all policy areas and provides an overview of the effects, the risks and the uncertainties attributable to the interventions. By quantifying the extent of possible advantages and disadvantages and to value them (expressed in Euros), the MKBA provides insight into the impact of the intervention on social welfare [30]. In other words: the balance of benefits measured in Euros minus costs. This balance also includes the costs and consequences of the intervention on components of the social wealth for which no market prices exist (e.g., security, culture and social cohesion) [31]. By monetizing the impact of the interventions they become comparable. Furthermore, answers can be provided to the question of whether the social costs of a measure outweigh the social benefits. Of course it is also possible that knowledge regarding the consequences is too uncertain to quantify in monetary terms. Nevertheless, MKBAs can be useful to display, in an orderly way, what is known about the intervention, and what is not. The decision-making can then be structured and become pragmatic in its approach. In summary, an MKBA can be used to steer future practice by providing insight into the potential changed overall social welfare and the distribution of costs and bene-

fits of the diverse range of employability interventions for employees with health problems.

To further aid decision-making in practice, insight into expected long-term costs and consequences should be valued too. As mentioned earlier, an 18-month follow-up period would be preferred to assess the full impact of an intervention. However, because of several practical limitations none of the studies were able to achieve this. The six-month follow-up period used throughout this PhD-study can therefore be insufficient to capture the true consequences of the analyzed interventions. Although the data collection methods chosen were suitable and as complete as possible for conducting economic evaluations, the measurement is limited to the length of the trial to which the economic evaluation is attached. In future research, it would be innovative and interesting to account for the costs and consequences after this period. Economic evaluation using decision analytical modeling can bridge the gap between trial-based observations and the expected long-term costs and consequences. By using decision analytical modeling, the expected costs and consequences of decision options regarding the implementation of employability interventions could be compared by synthesizing information from multiple sources and applying mathematical techniques [32]. The aim would be to provide decision makers and relevant stakeholders with the best available long-term evidence to reach a decision by comparing alternative options.

Lastly, it is hypothesized that employees are not used to the 'treating' role of occupational physicians such as offered in several employability interventions. Within the Dutch legislation, treatment and sickness certification are separate tasks, performed by different professionals without exchange of medical information unless the patient/employee gives consent [33]. In the studies included in this thesis, employees could have felt uncomfortable meeting with employers, healthcare providers and occupational physicians all at once, or within a very short time span. In an attempt to integrate occupational health more in the 'treating' healthcare system, collaboration between the different stakeholders (e.g. the occupational physician, the employee, the employer, etc.) needs to be improved. Although it is an assumption that employees feel uncomfortable with the 'treating role' of the occupational physician, the hypothesis is interesting to study, for example with qualitative research methods. Additionally, awareness of worker preferences should be assessed in order for employability interventions to embrace a more individualized approach, which meets the preferences of workers. Discrete-choice experiments are commonly known preference-based measures in economic evaluation [34], which can be used to comprehend employee's preferences regarding employability interventions.

Concluding remarks

Occupational health is a complex field of study and practice because of its transdisciplinary nature with multiple stakeholders involved. Many employability interventions exist that aim for sustainable employability among employees with health constraints. Nonetheless, appropriate decision making for implementation might be hampered for many reasons (such as methodological reasons). Economic evaluation results can aid to inform these decisions. The findings of this PhD-thesis could be used to guide decision making by outlining solutions to various methodological challenges, and by deducing the generated economic evaluations results of employability interventions assessed in practice. The information as presented throughout this PhD-thesis can support and guide diverse stakeholders such as researchers, occupational health practitioners, employees, employers, and decision makers, in selecting the employability intervention to be assessed through research or to be implemented in practice. The findings from this PhD-study help to substantiate these decisions. The included studies showed that the field of occupational health is not that easily affected, despite of the relative positive economic evaluation results of the employability interventions. A clear interface between what stakeholders find relevant, insight in context-specific characteristics of an intervention, and adequate measurement and valuation tools for both costs and outcomes is warranted.

References

1. Khan KS, Kunz R, Kleijnen J, Antes G. Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine*. 2003 Mar;96(3):118-21.
2. Noben CY, Evers SM, Nijhuis FJ, de Rijk AE. Quality appraisal of generic self-reported instruments measuring health-related productivity changes: a systematic review. *BMC Public Health*. 2014;14:115.
3. Mokkink LB, Terwee CB, Knol DL, Stratfor, PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010;10:22.
4. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2011 Jul 6.
5. Noben CY, Nijhuis FJ, de Rijk AE, Evers SM. Design of a trial-based economic evaluation on the cost-effectiveness of employability interventions among work disabled employees or employees at risk of work disability: The CASE-study. *BMC Public Health*. 2012 Jan 18;12(1):43.
6. Noben C, Smit F, Nieuwenhuijsen K, Ketelaar S, Gartner F, Boon B, et al. Comparative cost-effectiveness of two interventions to promote work functioning by targeting mental health complaints among nurses: pragmatic cluster randomised trial. *Int J Nurs Stud*. 2014 Oct;51(10):1321-31.
7. Drummond M, Manca A, Sculpher M. Increasing the generalizability of economic evaluations: recommendations for the design, analysis, and reporting of studies. *Int J Technol Assess Health Care*. 2005 Spring;21(2):165-71.
8. Glick H, Doshi G, Sonnad SS, Polsky D. *Economic evaluations in clinical trials* New York: Oxford University Press; 2007.
9. Byford S, Raftery J. Perspectives in economic evaluation. *BMJ*. 1998 May 16;316(7143):1529-30.
10. Mattke S, Balakrishnan A, Bergamo G, Newberry SJ. A review of methods to measure health-related productivity loss. *Am J Manag Care*. 2007 Apr;13(4):211-7.
11. Uegaki K, de Bruijne MC, Anema JR, van der Beek AJ, van Tulder MW, van Mechelen W. Consensus-based findings and recommendations for estimating the costs of health-related productivity loss from a company's perspective. *Scand J Work Environ Health*. 2007 Apr;33(2):122-30.
12. Knies S, Severens JL, Ament AJ, Evers SM. The transferability of valuing lost productivity across jurisdictions. differences between national pharmacoeconomic guidelines. *Value Health*. 2010 Aug;13(5):519-27.
13. Oostenbrink JB, Koopmanschap MA, Rutten FF. Standardisation of costs: the Dutch Manual for Costing in economic evaluations. *Pharmacoeconomics*. 2002;20(7):443-54.
14. Smith KA, Rudmik L. Cost collection and analysis for health economic evaluation. *Otolaryngol Head Neck Surg*. 2013 Aug;149(2):192-9.
15. Tompa E, Culyer AJ, Dolinschi R. *Economic Evaluation of Interventions for Occupational Health and Safety: Developing Good Practice*. New York: Oxford University Press; 2008.
16. Koopmanschap M, Burdorf A, Lötters F. *Work absenteeism and productivity loss at work*. Loisel P, Anema JR, editors: Springer; 2013. 513 p.
17. van den Hout WB. The value of productivity: human-capital versus friction-cost method. *Ann Rheum Dis*. 2010 Jan;69 Suppl 1:i89-91.
18. Krol M, Brouwer WB, Severens JL, Kaper J, Evers SM. Productivity cost calculations in health economic evaluations: correcting for compensation mechanisms and multiplier effects. *Soc Sci Med*. 2012 Dec;75(11):1981-8.
19. Evers HGJ. *The Improved Gatekeeper Act. A practical guideline* [In Dutch: De Wet Verbetering Poortwachter. Handleiding voor de praktijk]. Deventer: Kluwer Juridisch; 2007.

20. Hakkaart L, Tan SS, Bouwmans, CAM. Manual for cost research. Methods and standards costs for economic evaluations in healthcare [in Dutch: Handleiding voor kostenonderzoek. Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg]. Rotterdam, the Netherlands: Instituut voor Medical Technology Assessment, Erasmus Universiteit Rotterdam, 2010.
21. Sassi F. Calculating QALYs, comparing QALY and DALY calculations. *Health policy and planning*. 2006 Sep;21(5):402-8.
22. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ*. 2005 May;14(5):487-96.
23. Krol M, Brouwer W, Sendi P. Productivity costs in health-state valuations : does explicit instruction matter? *Pharmacoeconomics*. 2006;24(4):401-14..
24. Drummond M, Coyle D. The role of pilot studies in the economic evaluation of health technologies. *Int J Technol Assess Health Care*. 1998 Summer;14(3):405-18.
25. Arain M, Campbell MJ, Cooper CL, Lancaster GA. What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC medical research methodology*. 2010;10:67.
26. Thabane L, Ma J, Chu R, Cheng J, Ismaila A, Rios,LP, et al. A tutorial on pilot studies: the what, why and how. *BMC Med Res Methodol*. 2010;10:1.
27. Lofland JH, Pizzi L, Frick KD. A review of health-related workplace productivity loss instruments. *Pharmacoeconomics*. 2004;22(3):165-84.
28. Barber JA, Thompson SG. Analysis of cost data in randomized trials: an application of the non-parametric bootstrap. *Stat Med*. 2000 Dec 15;19(23):3219-36.
29. Brouwer WB, Koopmanschap MA, Rutten FF. Productivity costs in cost-effectiveness analysis: numerator or denominator: a further discussion. *Health Econ*. 1997 Sep-Oct;6(5):511-4.
30. Romijn G, Renes G. General guidance for societal cost benefit analysis. [in Dutch: Algemene leidraad voor maatschappelijke kosten baten analyse.]: Planbureau voor de Leefomgeving 2013. Available from: <http://www.pbl.nl/publicaties/algemene-leidraad-voor-maatschappelijke-kosten-baten-analyse>.
31. Drost RMWA, Paulus ATG, Ruwaard D, Evers SMAA. Manual for inter-sectoral costs and benefits of (preventive) interventions: Classification, identification and cost prices. [in Dutch: Handleiding intersectorale kosten en baten van (preventieve) interventies: classificatie, identificatie en kostprijzen.] Maastricht Maastricht University, Department of Health Services Research, 2014.
32. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. *Value Health*. 2003 Jan-Feb;6(1):9-17.
33. Elders L, Burdorf A. Occupational medicine in The Netherlands. *Occup Med (Lond)*. 2010 Jun;60(4):314.
34. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Appl Health Econ Health Policy*. 2003;2(1):55-64.

CHAPTER 12

Valorization of research results

Work is central in people's lives. Work influences well-being and impacts the physical, mental, financial, and social health of individuals. The health and growth of our communities is thereby directly influenced by our ability to maintain a healthy workforce. Hence, it is important that those employees who get influenced in their work-ability because of health complaints can maintain their place in the workforce. This valorization chapter addresses the main findings of this PhD-research relevant to the most prominent stakeholders present in the workforce, and thereby in occupational health. The sections below describe the target groups and elaborate on various points of interest which have the potential to be valorized and appear relevant for each stakeholder group.

Employees with health complaints

The employee is seen as the individual who is absent from the workplace due to health complaints or who is experiencing substantial burden of the health complaints while at work, hampering his or her employability.

Message from the thesis

Changed productivity is an important asset of employability interventions and should therefore be taken into account when assessing the cost effectiveness of employability interventions. For employees, however, increased productivity might not always be the primary aim, and thus not the main outcome of interest in the analyses. Individual employees are often more interested in other benefits associated with sustainable employability such as quality of life, work ability, self-efficacy, and so on.

Points of interest

Before selecting or implementing an employability intervention, it is important to ensure that the outcomes are also important from an employee perspective. Measuring a broader range of outcomes within the analysis of employability interventions will provide outcome evidence to select the most adequate employability interventions. Employees with health complaints need to be aware of their current and future (financial) position, as well as their current and future health state. Both employees' needs and preferences require visibility. Any assistance from professionals, researchers, or experts needs to be put in practice.

Efforts made to disseminate this knowledge

The following efforts are made to disseminate the knowledge: publication of study results of diverse employability interventions in scientific journals; publication of diverse methodological investigations in scientific journals; and several press releases and other publication types (such as reports).

Additional actions to be taken

Although the importance of outcome measures such as quality of life, productivity, and work ability are unquestionably relevant, more insight in what is relevant and important for the employee is needed. Identifying a broader range of employee-centred outcomes, as well as employee preferences, might indicate more properly whether the employee's goals fit the aims of the other stakeholders. Gaining these insights might indicate why certain employability interventions are more, or less, cost effective when compared to common practice.

Co-Workers

This interest group relates to those individuals directly affected by the work ability of the individual employee with health complaints.

Message from the thesis

Measuring the consequences of an employee's work-related health problems for the co-workers can be incorporated in measuring the so-called 'compensation mechanisms'. Compensating for the lost productivity will influence the related costs. Co-workers can take over (and thus compensate for) the lost work of the employee with work-related health complaints. The impact thereof is illustrated in the changed productivity costs, assuming less productivity costs if the work is compensated for by co-workers.

Points of interest

When measuring changed productivity due to work-related health complaints, it is recommended to also include the measurement of compensation mechanisms (e.g. co-worker compensation). Co-worker's efforts need to be recognized.

Efforts made to disseminate this knowledge

The following efforts are made to disseminate the knowledge: publication of study results of diverse employability interventions in scientific journals; publication of diverse methodological investigations in scientific journals; presentations and workshops at (inter)national conferences; and several other publication types (such as reports).

Additional actions to be taken

The consequences for co-workers need to be taken into account to accommodate sustainable employability for employees with work-related health complaints, and to prevent the co-workers from developing work-related health complaints themselves. Insight is required in both the burden of extra work experienced by the co-worker when compensating for the employee with work-related health complaints,

and the co-worker's expectancies towards the returning worker. Researchers have a large role in gaining and disseminating these insights.

Employers

The group of employers addressed are those organisations which employ workers or those individuals that relate to the employee via the workplace. One can think of company owners, supervisors, human resource managers, and so on.

Message from the thesis

The employability interventions assessed from an employer's perspective have demonstrated that it is recommended to invest in employability because productivity increased in a cost-beneficial way. For employees in the Netherlands these analyses are especially relevant since they are obliged under the Dutch law to guarantee at least 70% of the salary of their absent employees with health complaints. If employers are about to invest in employability interventions, it is essential to provide them with this information so they can make evidence-informed decisions.

Points of interest

Aside from increasing productivity, employers are essential actors to also decrease the burden of work-related health complaints. A distinction is preferably to be made between those stakeholders which have to oversee the broader vision of the company - and thus primarily aim for increased productivity- and those in direct contact with the employees, who aim for reduced burden of work-related health complaints.

Efforts made to disseminate this knowledge

The following efforts are made to disseminate the knowledge: contacts with (a group of) employers to discuss the needs and to disseminate the newly gained knowledge; publication of study results of diverse employability interventions in scientific journals; publication of diverse methodological investigations in scientific journals; presentations and workshops at (inter)national conferences; and several press releases and other types of publications (such as reports)

Additional actions to be taken

For company owners and human resource managers it is necessary to gain insight into the impact of the work-related health complaints on the functioning of their organisation. Assessing the cost effectiveness of employability interventions provides an excellent method to gain more insight in impending efforts to control the costs of changed productivity. Direct supervisors play a key-role in sensing what happens at the workplace. Identifying a broader range of employee-centred outcomes, as well as employee preferences, might indicate more properly whether the

employee's goals fit the aims of the direct supervisors. Gaining these insights might indicate why certain employability interventions are more, or less, cost effective when compared to common practice.

Insurance companies

Depending on the country's social welfare system, insurance companies become the ones paying for activities designed to help employees recover from their health complaints and assist them towards employability, or to compensate for their work disability. Based on the Dutch social welfare system, insurance companies can be involved as income insurer for companies (to insure the extended wage-payments in case of sickness absence), or as collective health insurer.

Message from the thesis

In the Netherlands, insurers for companies benefit from sustainable employability through lower or no 'benefits' or 'compensation payments' for lost productivity and the reduced extended wage-payments. Health care insurers also benefit from investing in employability interventions when less healthcare is utilized. In all, investing in cost effective employability interventions helps employees to maintain 'financially healthy'. Sustainability is an important aspect as it demonstrates successful employability and reduces the negative impact, and lowers the likelihood, of recurrent absenteeism and presenteeism.

Points of interest

Insurers can have varying interest as they might concentrate on healthcare costs associated with work-related health complaints, lost productivity costs due to absenteeism and presenteeism or the funding of (a potential range of) employability interventions. Despite of these varying rationales, and perhaps divergent motives for payments, it is the insurer's interest to invest, or at least facilitate, sustainable employability.

Efforts made to disseminate this knowledge

The following efforts are made to disseminate the knowledge: publication of study results of diverse employability interventions in scientific journals; publication of diverse methodological investigations in scientific journals; presentations and workshops at (inter)national conferences; provided input for symposia and conferences addressing insurance-related issues; and input for press releases and reports to provide guidance for insurance-related issues

Additional actions to be taken

Different motives for payments from insurers are best to be supported by clear, concise and transparent research evidence in order for insurance companies to

make evidence-informed decisions to invest in employability interventions. The subjective weights attached to insurer's motives are best prevented by systematically providing evidence on the effects, both in terms of economic motives (e.g. productivity costs, healthcare utilization costs), and in terms of more 'humanitarian' motives such as quality of life, preferences, work ability, and so on. These insights might also more properly indicate why certain employability interventions are more, or less, cost effective when compared to common practice.

Healthcare providers

This group of stakeholders includes all those that provide assistance in helping the employee with health complaints to recover and sustainably (return to) work. Although the main objective of healthcare providers is focused on optimising 'health', healthcare providers become increasingly involved in occupational healthcare as well.

Message from the thesis

Healthcare providers play an important role in preparing an employee with work-related health complaints to return to work. Furthermore, they play an essential role in work-disability prevention. Healthy and sustainable employability can be seen as a function of (medical) treatments whereby patients return to active participation.

Points of interest

As healthcare providers consist out of a very heterogeneous group (such as general practitioners, specialists, and so on) and employees with work-related health complaints can consult a diverse range of healthcare providers, a clear role description for each healthcare provider is preferred. Sickness severity and duration, as well as the context of both the condition and labour participation need to be acknowledged by the different healthcare providers. In order to reduce a potential negative impact of the variation of what is priority is based on a clear description of the healthcare provider's role and tasks, and on cooperation and communication between these stakeholders.

Efforts made to disseminate this knowledge

The following efforts are made to disseminate the knowledge: contacts with (groups of) healthcare providers to discuss the needs and to disseminate the newly gained knowledge; publication of study results of diverse employability interventions in scientific journals; publication of diverse methodological investigations in scientific journals; presentations and workshops at (inter)national conferences; and several press releases and other publication types (such as reports).

Additional actions to be taken

More guidance and insight might help those healthcare providers that experience difficulties in accurately reflecting on an individual's work disability status. Open communication between the healthcare providers and patients with paid work regarding the patient's work, work-ability and preferences should be a standard point of interest.

Reintegration professionals & agencies

This group of stakeholders includes those that work in the reintegration field. In the Netherlands, this field has evolved into a mix of public and private organizations, with a wide range of reintegration professionals. For employees, the reintegration professionals most commonly known are 'occupational physicians'. In the Netherlands, the employers hire occupational physicians. They deliver a private service, but work under public legislation. According to the Dutch law, occupational physicians have a legal obligation to offer guidance to employees with health complaints in the first two years of sickness absence. Additionally, employers can also contract a private reintegration agency to offer for example more extensive training and counseling.

Message from the thesis

Just as wide and diverse as the involvement of diverse reintegration professionals or agencies can be, as broad are their approaches. Insight in the cost effectiveness of employability interventions provided by reintegration professionals and reintegration agencies is needed to winnow the wheat from the chaff.

Points of interest

In practice, not all reintegration agencies operate in a systematic manner. Evidence-based employability interventions are seldom put in practice; it is often more a matter of availability. Systematic intervention-mapping is rarely done. Lacking insights in the cost effectiveness of employability interventions hamper their continuation and optimization.

Efforts made to disseminate this knowledge

The following efforts are made to disseminate the knowledge: publication of study results of diverse employability interventions in scientific journals; publication of diverse methodological investigations in scientific journals; presentations and workshops at (inter)national conferences; and several press releases and other publication types (such as reports)

Additional actions to be taken

More research is needed on whether and why reintegration interventions are successful or not. More evidence-based employability interventions are needed. By conducting economic evaluations, not only insight in the worthiness of investing in employability interventions, but also insight in the feasibility of these interventions is gained. Only by more systematic assessments, employability interventions can be compared, adjusted and potentially put in practice. Researchers and reintegration professionals-, and agencies should work more closely together.

Policy makers

This interest group is located at the local and/or national governmental level, attempting to optimise occupational health and to reduce social welfare costs. Policy makers are defined in terms of their role in maintaining and regulating society.

Message from the thesis

As policy makers are (legally) responsible to enhance labour participation, and to reduce work-related health problems, economic evaluation results from the societal perspective are preferably to be used to aid evidence-informed policy making.

Points of interest

It could be argued that the view of policy makers is embodied in societal legislations, approaches to social security, and healthcare service provisions. Despite of the relatively positive economic evaluation results of the employability interventions assessed in this thesis, the field of occupational health requires more evidence to inform policy-making. Policy makers need to shape both the macro level of employability (e.g. institutions), as well as the context wherein work-related health problems occur (micro level). The motivation to strive for sustainable employability might rest on different foundations (e.g. more seen as a social, or more as an individual responsibility).

Efforts made to disseminate this knowledge

The following efforts are made to disseminate the knowledge: publication of study results of diverse employability interventions in scientific journals; publication of diverse methodological investigations in scientific journals; presentations and workshops at (inter)national conferences; and several press releases.

Additional actions to be taken

Policy-makers would benefit from acquiring the knowledge of key methodological principles underlying economic evaluations. Exchange and access of information from health technology assessment researchers or health economists for policy-

makers would be helpful. At the moment the exchange of information, and thus information-based policy, is already commonly used within the field of high value cost conscience healthcare and within the assessment of medicines (e.g. structural guideline revisions). Within occupational health it is rather lacking. More insight in methodological and content-related indications, and more insight in economic evaluation results, could aid policy-makers to ensure that occupational health programs are providing mandated and high quality employability interventions, and thereby maintaining and regulating society.

Summary

Health problems continue to limit labor participation, creating a growing need for sustainable employability. Structural methods are needed to develop appropriate responses when health problems impede employability and labor participation. This PhD-thesis addresses this by providing insight how the methodology underlying economic evaluations within occupational health care can be improved. The first central aim throughout this PhD-study is therefore to *generate knowledge and insight into the methodological challenges when conducting economic evaluations of employability interventions* (Part I). Given the results, efforts to improve the current methodology are required. Throughout this PhD-study, the required efforts were applied to practice in order to conduct methodological high quality economic evaluations in employability interventions. To properly assess the added value of both existing-, and innovative employability interventions, this PhD-thesis furthermore describes the underlying health economic aspects, and the economic evaluation results of employability interventions aimed at achieving sustainable employability for employees with health problems. The second central objective of this PhD-thesis is therefore the application of the generated methodological knowledge to different employability interventions, and the thereby *generated economic evaluation results of different existing and innovative interventions aimed at achieving sustainable employability for employees with health problems* (Part II and III).

Part I: Methodological Challenges

Chapter 2 presents a systematic review in which the measurement properties of generic self-reported instruments that measure health-related productivity changes are critically appraised. 25 articles assessing the reliability, validity and responsiveness of 15 different generic self-reported instruments measuring health-related productivity changes were included. The Work Limitation Questionnaire (WLQ) was most frequently evaluated and yielded moderate to strong positive evidence for content and structural validity. The Stanford Presenteeism Scale (SPS) and the Productivity and Disease Questionnaire (PRODISQ) also yielded strong positive evidence for several measurement properties. The overall evidence on the psychometric properties of the other 12 instruments was limited and of fair to poor quality. Based on these findings, the WLQ, SPS, and in a Dutch context, the PRODISQ are cautiously preferred until high-quality studies are in place to accurately assess the measurement properties of the currently available instruments. This study also warrants careful consideration of the content of the instrument, usage purpose,

target country and population of interest, and available evidence to guide decision making.

The results of another systematic review are presented in **Chapter 3**. The exchangeability of self-reported and administrative healthcare resource use measurements for cost estimation was studied. The methodological reporting quality of studies comparing the validation evidence of healthcare resource use measurement tools was assessed with an appraisal tool. This tool was derived by combining aspects from the Guidelines for Reporting Reliability and Agreement Studies and The Standards for Reporting Diagnostic Accuracy. Out of the 16 included studies, seven scored 'good' on more than 75% of the reporting criteria. In the end, six studies scored 'good' on the minimal criteria for reporting. These criteria included: (i) a clear description of the measurement methods with a critical review of evidence relevant for the study, (ii) a rationale for the relationship between the self-report and the administrative data, (iii) the methods or measurement procedures for the self-report, and (iv) the methods and measurement procedures for the administrative data. Despite varying levels of agreement among the different data sources, the validation evidence of the small number of studies with adequate methodological reporting quality cautiously supported the exchangeability of both the self-reported and administrative resource use measurement methods.

Chapter 4 describes how to preferably conduct economic evaluations of multiple different employability interventions based on a similar methodological framework. This framework enables the development of guidelines on how to conduct economic evaluation in this field and thereby contributes to the decision-making regarding treatment options in occupational health practice. The paper describes a consistent and transparent methodological design to analyze the cost effectiveness of five different employability interventions directed at work disabled employees with divergent health complaints. The comparisons between the employability interventions and usual care or the control conditions were pragmatic in nature replicating the actual ((occupational) health care) setting in which the studies occur. The design is compliant for different settings, for example when different unit prices for resource utilization, or divergent outcome measures, are preferred. According to sample-size calculations, each employability intervention needed a total of 142 participants to detect decreased sick leave duration. Eligible participants were employees aged 18 to 63, working at least 12 hours per week, and at risk of work disability, or already work-disabled due to medical restrictions. The economic evaluation results for each employability intervention are described elsewhere. This chapter discussed the methods, strengths and weaknesses of the study protocol.

Given the results of the studies included in Part I of this PhD-thesis, the economic evaluation of existing- and new employability interventions are based on solid methodologies, accounting for stakeholder interests, and proper methodological reporting quality to ensure transparency. The economic evaluation results are summarized in the paragraphs below.

Part II: Existing practices

In **Chapter 5**, the results of a comparative cost effectiveness of two strategies to promote work functioning among nurses by reducing symptoms of mental health complaints are presented. In total, 617 nurses from one academic medical centre in the Netherlands were included and divided over either the control condition (online screening for mental health problems without feedback about the screening results), the occupational physician condition (screening, feedback and referral to the occupational physician for screen-positive nurses), or the e-mental health condition (screening, feedback, and referral to e-mental health). After six months, significant improvement in work functioning occurred in all three groups. The incremental cost effectiveness ratio for the occupational physician condition versus the control condition was dominant, suggesting cost savings of €5,049 per treatment responder. The occupational physician condition resulted in greater treatment responses for less costs relative to the control condition and was therefore recommended to be put in practice. The incremental cost effectiveness ratio for the e-mental health condition versus the control condition was estimated at €4,054 (added costs) per treatment responder, producing less treatment response than the control condition and therefore not recommended to be put in practice.

Chapter 6 addresses the question to whether a preventive intervention in the work setting is cost-saving from a business perspective. Alongside a cluster-randomized trial in a Dutch academic hospital, the control condition (screening without feedback) was compared to the experimental condition (screening, personalized feedback and referral to the occupational physician for screen-positive nurses) to evaluate the balance between the invested intervention costs and the cost offsets stemming from improved productivity. Net-savings of €244 per nurse were observed when only absenteeism is regarded and €651 when presenteeism is also taken into account. Within half a year, the costs of offering the preventive intervention were more than recouped. A return-on-investment ranging from €5 up to €11 per Euro invested was found. Offering the preventive intervention represents a favourable business case as seen from the employer's perspective.

Chapter 7 focusses on the results of a pilot cost effectiveness study. This study aimed to detect how to improve a web-based employability intervention aimed at increasing self-perceived work ability among employees who experienced problems in relation to their employability. This study detected elements requiring improvements and thereby demonstrated the value of cost effectiveness analyses within a pilot study, preceding a full-scale study. First, content related improvements were recommended based on the low uptake and compliance rates, particularly for the online intervention element. Additional analyses concluded that the intervention appeared too general for specific situations, such as prolonged health problems requiring specific questions on work and health. Furthermore, the web-based employability intervention was not associated with an economically relevant impact over regular trade union support for a heterogeneous population with

work- and disability related concerns. However, there was a tendency females and participants aged 50 or older might be more sensitive for an intervention effect. Ongoing tailoring and better targeting was therefore recommended. Secondly, improvements regarding the study design of the web-based employability intervention were also recommended based on the results of the pilot economic evaluation. Negligible incremental differences on work ability and quality adjusted life years were unable to demonstrate an effect of the web-based intervention. Future studies should avoid discussion on whether the most appropriate outcome was selected to assess the true effect of the intervention, or whether the effect of the intervention was influenced by other factors (for instance derived information other than via the web-based employability intervention). In addition, the relative contribution of patient- and family out-of-pocket costs to the total overall costs was minimal. Data gathering for these cost categories could be omitted to minimize the burden for the employees and their environment. Based on the pilot economic evaluation results, the intervention content needs to be improved with respect to targeting and intensity. Regarding the study design, reconsideration of the selected outcome measures and the collection of cost data is preferred.

Part III: Innovative interventions

In the third part, in **Chapter 8** a field study is described that assessed the cost effectiveness, -utility, and -benefit of a new organizational return-to-work (RTW) intervention to improve Cooperation between Sick-listed employees and their Supervisors (COSS). After six months follow-up, employees in the COSS-group reduced almost 16 calendar days of absenteeism and saved approximately €27 per day of work resumption. The cost utility analysis also exhibited reduced costs; however the gain in quality adjusted life years was uncertain. From an organizational perspective, the net benefits of COSS versus common practice yielded a productivity gain of almost €396. COSS has the potential to improve productivity at reduced societal and organizational costs when implemented among sick-listed employees. Long-term research needs to detect whether COSS also has the potential reaching sustainable RTW.

Chapter 9 provides the results of a short-term economic evaluation. First insights into the cost effectiveness of an intervention coordinating clinical- and occupational care at the worksite are provided. This intervention is aimed to support the work participation of workers with rheumatoid arthritis. After six months, both at-work productivity loss and QALY differences were in favor of the control group. Together with the higher costs of the intervention group compared to the control group, no evidence was found to support the implementation of the intervention. The main cost effectiveness and cost utility analyses show that the intervention was more expensive and less effective than the control condition. Sensitivity analyses supported these findings.

In **Chapter 10**, another economic evaluation was conducted. A tailored educational training program for general practitioners was provided to increase GPs' awareness of work-related health problems. This training program aimed to improve treatment strategies in the primary care setting. Awareness for work-related health problems in the GP practice is assumed to increase work-related self-efficacy and quality of life for patients in paid work. After six months follow-up, the differences in the main outcome measures between the intervention group and the usual GP care group were minimal. A negative incremental cost effectiveness ratio revealed fewer effects at higher costs for both the cost effectiveness and -utility analyses. Based on cost-related grounds the intervention could therefore not be recommended for implementation.

Part IV: Discussion

In **Chapter 11** the main practical and methodological considerations and considerations regarding the employability interventions that aim for sustainable employability among employees with health constraints are touched upon.

Practical considerations regard the recruitment of organizations and participants to take part in the research. To conduct an economic evaluation, at least a minimum required number of participants, a comparative design, and a comparable control group needs to be guaranteed. These practical considerations might slow down data collection as they require contacting other parties. Another practical consideration is the follow-up time, which is preferably stated before the study begins, but often not achieved due to practical issues.

Methodological considerations concentrate around the perspective of study. Within this field of study the perspective often has to deviate from the preferred societal perspective according to the health economic guidelines when accounting for specific stakeholder's interest and when preventing the risk of double counting productivity changes. Further methodological considerations concern the methodological (reporting) quality of measurement tools, more specifically those that measure and value costs. Uniform and transparent methods are still lacking. Next, generalizability issues might also be at stake when applying these findings to another country, and thus another social security system. A final methodological consideration distresses the lack of insight into changes in the quality of life due to health problems impeding labor participation, and how employability interventions might improve the effect of productivity on health and quality of life.

Final considerations regard the intervention. The studied control groups were already likely effective, limiting the visibility of the effectiveness of different employability interventions. In addition, some effect measurements did not represent what is truly relevant to expect from an employability intervention for employees with health problems (e.g. generic QALYs). The results of the included studies demonstrated that the field of (Dutch) occupational health is not that easily affected, even when the economic evaluation results of the employability interventions

SUMMARY

are relatively positive. Adequate implementation of the employability intervention in practice requires meeting the needs of all stakeholders. Hence, decision making also requires the acknowledgement of context-specific characteristics, which make the employability intervention more or less relevant to particular stakeholders. Pilot economic evaluations, conducted before large-scale studies, appeared to provide valuable insight into how and where to target the intervention, given its population of interest.

In future research more attention could preferably be paid to subgroup analyses to account for participant heterogeneity, which provides valuable information on how to target the interventions aimed at achieving sustainable employability for employees with health problems. In addition more focus on characterizing the uncertainty around the joint cost-effect estimates (CEA-plane) is preferred. At the moment no visual methods, such as a uniform willingness-to-pay or willingness-to-accept for increased health-related productivity, has yet been developed to help occupational health decision makers interpret the economic evaluation results. Also, more insight into health-related productivity and research on factors influencing changed productivity (e.g. compensation) and the measurement of productivity costs would be helpful to interpret the results. Portraying and valuing the advantages and disadvantages of an employability intervention (from the societal perspective) will support the policy choices regarding the implementation of such an intervention. The MKBA is suggested as an appropriate method since it is applicable for all types of interventions in all policy areas and provides an overview of the effects, the risks and the uncertainties attributable to the interventions.

Samenvatting

Recentelijk zijn diverse interventies ontwikkeld die streven naar duurzame inzetbaarheid van werknemers met gezondheidsproblemen. Deze interventies worden 'inzetbaarheidsinterventies' genoemd. Gezien de betrokkenheid van meerdere belanghebbenden is gezondheid op het werk een complex onderzoeksgebied. Hierdoor kunnen besluiten over het inzetten van de juiste inzetbaarheidsinterventies bemoeilijkt worden. Verschillende afwegingen moeten worden gemaakt. Onderzoek, en voornamelijk economisch evaluatieonderzoek, kan helpen om deze afwegingen op een gegronde manier te maken. De resultaten ondersteunen besluitvorming over het al dan niet implementeren van de specifieke inzetbaarheidsinterventie en informeren onder welke voorwaarden dit gedaan moet worden om het beoogde effect te bereiken.

Deze dissertatie beoogt: (i) het genereren van *kennis en inzicht in de methodologie* van economisch evaluatie onderzoek bij inzetbaarheidsinterventies (deel I), en, (ii) het interpreteren van de resultaten van *economisch evaluatie onderzoek van zowel bestaande als innovatieve inzetbaarheidsinterventies* gericht op het realiseren van duurzame inzetbaarheid van werkenden met gezondheidsproblemen (deel II & III).

Deel I: Methodologische uitdagingen

In **hoofdstuk 2** werden generieke, zelf-gerapporteerde instrumenten die gezondheidsgerelateerde productiviteit meten kritisch beoordeeld in een systematische review. In totaal konden 15 verschillende instrumenten geïncludeerd en geëvalueerd worden. De betrouwbaarheid, validiteit en responsiviteit van deze verschillende instrumenten werd onderzocht. The Work Limitations Questionnaire (WLQ) was het meest frequent geëvalueerd en die studies leverden een matig tot sterk positief bewijs voor zowel inhoudelijke als structurele validiteit. Ook de Stanford Presenteeism Scale (SPS) en de PROductivity and DISease Questionnaire (PRODISQ) leverden sterk positief bewijs voor meerdere psychometrische eigenschappen. Op basis van deze bevindingen werden de WLQ, SPS, en (in een Nederlandse context) de PRODISQ voorzichtig aanbevolen. Om de besluitvorming te begeleiden werd tevens aanbevolen zorgvuldige afwegingen te maken ten aanzien van de inhoud van het instrument, het streefdoel en de doelgroep.

In een tweede systematische review in **hoofdstuk 3** werden meetmethoden bestudeerd om de kosten van gezondheidszorg te meten. De 'uitwisselbaarheid' van zelf-gerapporteerde en administratieve data ten aanzien van het zorggebruik werd onderzocht. Om hierover uitspraken te doen, werden gepubliceerde validatie-

studies beoordeeld. Deze beoordeling richtte zich op de kwaliteit van de methodologische verslaglegging aan de hand van een nieuwe 'tool'. Deze tool werd samengesteld door aspecten van verschillende richtlijnen te combineren (*Guidelines for Reporting Reliability and Agreement Studies* en *The Standards for Reporting Diagnostic Accuracy*). Uit de resultaten bleek dat zelf-gerapporteerde als administratieve meetmethoden over het algemeen goed genoeg uitwisselbaar zijn.

Vervolgens werd in **hoofdstuk 4** beschreven hoe een consistente en transparante methodologische onderzoeksopzet ontwikkeld werd. Deze onderzoeksopzet beschrijft hoe de economische evaluatie van verschillende interventies gericht op duurzame inzetbaarheid bij werkenden met uiteenlopende gezondheidsklachten uit te voeren. Door interventies in het veld van arbeid en gezondheidszorg systematisch te onderzoeken wordt de vergelijking van deze behandelopties eenduidiger. De vergelijkingen tussen inzetbaarheidsinterventies en de controle condities (de gebruikelijke interventies) zijn bij voorkeur pragmatisch van aard om de werkelijkheid zo goed mogelijk te repliceren. De beschreven onderzoeksopzet kan breed ingezet worden en rekening houden met de vereisten van bijvoorbeeld diverse organisatie en aanbieders.

Deel II: Bestaande inzetbaarheidsinterventies

De eerste studie, beschreven in **hoofdstuk 5**, betrof een vergelijkende kosteneffectiviteitsanalyse van twee strategieën om het functioneren van verpleegkundigen te bevorderen door psychische symptomen te verminderen. In totaal werden 617 verpleegkundigen verdeeld over een controle conditie (online screening voor geestelijke gezondheidsproblemen zonder feedback over de resultaten van de screening), een 'bedrijfsarts-interventiegroep' (screening, feedback en doorverwijzing naar de bedrijfsarts voor screen-positieve verpleegkundigen), of een 'e-mental health groep' (screening, feedback, en verwijzing naar e-mental health). Gezien de dominante kostenbesparing van €5.049 ten opzichte van de controle conditie, werd implementatie aanbevolen voor de 'bedrijfsarts-interventie'. De incrementele kosteneffectiviteit ratio voor de 'e-mental health conditie' versus de controle conditie bedroeg €4.054 extra kosten. Implementatie van de 'e-mental health interventie' in de praktijk werd op basis van de vergelijking met de controle conditie niet aanbevolen.

In **hoofdstuk 6** werd vervolgens onderzocht of een preventieve interventie in dezelfde werksetting resulteerde in een positieve 'business case' vanuit het perspectief van de werkgever. De controle conditie (screening zonder feedback) werd vergeleken met de experimentele conditie (screening, persoonlijke feedback en doorverwijzing naar de bedrijfsarts voor screen-positieve verpleegkundigen). De interventie gerelateerde (investerings-)kosten werden afgezet tegen de kosten die voortvloeien uit veranderde productiviteit. Er werden netto besparingen variërend van €244 tot €650 per verpleegkundige waargenomen. Binnen een tijdsbestek van

een half jaar werd een positieve *return-on-investment* tussen €5 tot €11 per geïnvesteerde euro behaald.

Hoofdstuk 7 richtte zich op een pilot kosteneffectiviteit studie van een web-based inzetbaarheidsinterventie gericht op het vergroten van werkvermogen onder werknemers met gezondheidsklachten. De pilot kosteneffectiviteit studie onderzocht of (a) inhoudelijke verbeteringen, en/of (b) verbeteringen ten aanzien van de studieopzet nodig waren. Op basis van de lage respons en minimale gebruik van (onderdelen van) de interventie werden aanbevelingen gedaan om zowel de respons als het gebruik te bevorderen. Inhoudelijke aanbevelingen ten aanzien van de interventie werden gebaseerd op het gebrek aan noemenswaardige effecten van de interventie. Hoewel de verbeteringen ten aanzien van werkvermogen minimaal waren, detecteerden subgroep analyses groepen gevoeliger voor een effect van de interventie. Doelgerichter inzetten van interventieonderdelen voor specifiek gevoeliger groepen ('targetten') werd aanbevolen. Verbeteringen ten aanzien van de onderzoeksopzet werden ook aanbevolen op basis van de minimale effectiviteit van de interventie, vooral de marginale verschillen voor werkvermogen en de kwaliteit van leven. De relatieve bijdrage van kosten voor de patiënt en familie aan de totale kosten waren verwaarloosbaar, waardoor ook op basis van deze bevindingen aanbevelingen werden gedaan om verbeteringen door te voeren. Alvorens grootschalig onderzoek naar implementatie op te zetten, dragen pilot kosteneffectiviteit studies bij aan zowel inhoudelijke als methodologische inzichten om interventies in het veld van arbeid en gezondheidszorg te verbeteren.

Deel III: Innovatieve inzetbaarheidsinterventies

Hoofdstuk 8 beschreef de economische evaluatie van een innovatie gericht op de organisatorische aspecten die terugkeer naar arbeid beïnvloeden. De COSS-interventie richtte zich op de verbeterde samenwerking tussen verzuimende werknemers en hun werkgevers. Na zes maanden follow-up reduceerde verzuim (gemeten als initiële terugkeer naar werk) bij de werknemers in de interventiegroep met bijna 16 kalenderdagen en werd ongeveer €27 per dag bespaard door de werkherleving. Vanuit organisatorisch oogpunt leverde de interventie (vergeleken met de controle conditie) netto- productiviteitswinst op van bijna €396 per werknemer.

Hoofdstuk 9 presenteerde korte-termijn resultaten van een inzetbaarheidsinterventie die klinische- en arbeidsgerelateerde zorg trachtte te optimaliseren. De interventie ondersteunde de arbeidsparticipatie van werknemers met reumatoïde artritis. Na zes maanden interim analyse bleek dat de interventie duurder en minder effectief (in termen van arbeidsproductiviteit en kwaliteit van leven) was dan de controle conditie. Momenteel worden nog data verzameld voor een langere termijn follow-up.

In **hoofdstuk 10**, werd een op maat gemaakt educatief trainingsprogramma voor huisartsen economisch geëvalueerd. Het doel van de interventie was het bewustzijn van huisartsen ten aanzien van werk gerelateerde gezondheidsproblemen

in de huisartsenpraktijk te verhogen en hiermee de werk gerelateerde 'self-efficacy' en de kwaliteit van leven voor patiënten in betaald werk te verhogen. Na een interim-analyse van zes maanden bleken de verschillen in effectiviteit tussen de interventiegroep en de gebruikelijke huisartsenzorg-groep minimaal te zijn. Momenteel worden nog data verzameld voor een langere termijn follow-up.

Deel IV: Discussie

De bevindingen van deze dissertatie boden oplossingsrichtingen voor diverse methodologische uitdagingen binnen het complexe onderzoeksgebied van gezondheid op het werk. Daarnaast werden de resultaten van verschillende economische evaluaties bij inzetbaarheidsinterventie in de praktijk gepresenteerd en geïnterpreteerd. Hoewel de studies aantoonde dat de praktijk van arbeid en gezondheid zich niet zo gemakkelijk laat beïnvloeden, boden de (relatief positieve) economische evaluatie resultaten van de inzetbaarheidsinterventies toch diverse mogelijkheden voor de belanghebbenden. De resultaten van de studies beschreven in deze dissertatie kunnen onder andere onderzoekers, werknemers, werkgevers en beleidsmakers ondersteunen in het selecteren van inzetbaarheidsinterventies voor zowel verder onderzoek als implementatie in de praktijk. Een duidelijke uitwisseling tussen wat belanghebbenden relevant vinden, inzicht in context-specifieke kenmerken van een interventie, en adequate meting en waardering voor zowel kosten als uitkomsten, bleken essentieel.

Dankwoord

De reis is ten einde. En wat een reis was me dat! Als gevolg ligt hier nu dit proefschrift én ontstond er een 'nieuwe ik'. Ik heb enorm veel geleerd tijdens deze reis en besef maar al te goed dat ik deze eindbestemming nooit had kunnen bereiken zonder de hulp en steun van velen. Dit is de plek waarop ik jullie allen, mijn reisgezellen, uitgebreid wil bedanken.

Allereerst en in het bijzonder, alle deelnemers aan het onderzoek. Zonder jullie input was dit nooit gelukt. Jullie bijdrage heeft niet alleen dit proefschrift tot stand gebracht, maar draagt er ook toe bij dat het werk werkbaar wordt voor vele andere werknemers met gezondheidsklachten. Ik ben er zeker van dat ik ook in hun naam mag spreken wanneer ik zeg: "bedankt!"

Veel dank ook aan mijn promotor Prof. dr. Frans Nijhuis. Wat een enorm respect heb ik voor jouw optimistische kijk op de wereld. Daar kan ik nog veel van opsteken. Ook voor de manier waarop je met slechts een enkel woord tot anderen kan doordringen heb ik bewondering. Je bent gedurende dit traject een groot voorbeeld voor me geweest en kon heel goed aanvoelen wanneer ik 'gesust' moest worden. Ik ben je dan ook enorm dankbaar voor je vertrouwen in mij als onderzoeker. Ik zal altijd met veel waardering naar onze bijzondere samenwerking terugkijken.

Ook mijn promotor Prof. dr. Silvia Evers wil ik bedanken voor onze samenwerking. Silvia, je liet me altijd vrij om mijn eigen weg te zoeken maar stuurde het schip ook weer bij wanneer ik de foute koers dreigde te gaan varen. Ik ben dankbaar dat ik samen met jou deze tocht mocht maken en kijk er naar uit om de komende jaren verder samen te werken. Naast onze 'wetenschappelijke praatjes' was er ook tijd voor gezelligheid. Die combinatie heb ik altijd erg prettig gevonden, net zoals onze samenwerking in zowel onderzoek als onderwijs. Bedankt ook voor je 'open-mindedheid'. Daarzonder had dat sollicitatie Skype-gesprek vanuit Colombia naar Nederland waarschijnlijk nooit plaatsgevonden. Dankjewel voor alle kansen die je me geeft en gegeven hebt!

Dr. Angelique de Rijk, mijn copromotor. Ook jou wil ik van harte bedanken voor de samenwerking. Jouw kritische en analytische scherpte hebben mede mijn 'nieuwe ik' gevormd. Je hebt me geleerd altijd goed voor ogen te hebben waar ik heen wilde, welke input ik hiervoor nodig had en welke aanpak daar wellicht het best bij paste. Waar ik soms te zeer ging inzoomen op details zorgde jij er altijd voor dat ik nooit het grote geheel uit het oog verloor. Je hebt me altijd erg geholpen bij het structure-

DANKWOORD

ren en onderbouwen van bevindingen. Ik wil je dan ook bedanken voor de mogelijkheden die je me gegeven hebt om mij breder te ontwikkelen, zeker ook op het gebied van onderwijs.

Ik bedank ook graag de leden van de beoordelingscommissie voor het lezen en beoordelen van mijn thesis: Prof. dr. Andries de Grip, Prof. dr. IJmert Kant, Prof. dr. Werner Brouwer, Prof. dr. Carolyn Dewa, onder voorzitterschap van Prof. dr. Dirk Ruwaard.

Hiernaast wil ik ook graag mijn dankbaarheid tonen aan het Stichting Instituut Gak voor het financieel mogelijk maken van dit onderzoek. In het bijzonder veel dank aan Dr. Boudien Krol en Prof. dr. Peter van Lieshout voor hun interesse en advies.

Heel graag wil ik de collegae bedanken waarmee ik heb samengewerkt om de studies binnen dit onderzoekstraject tot een goed einde te brengen. Ik ben erg dankbaar dat ik betrokken mocht zijn bij zoveel projecten en zoveel nieuwe mensen heb leren kennen. Jullie hebben mij allemaal binnen jullie eigen mogelijkheden geholpen om dit resultaat te verwezenlijken. Ik hoop van harte dat we in latere stadia nog mogen samenwerken.

De afdeling 'Work and Social Psychology' (Maastricht University): Dr. Gemma van Ruitenbeek en Prof. dr. Fred Zijlstra, bedankt voor het vertrouwen en de voortvarende samenwerking tijdens de economische evaluatie van jullie project. Jullie inzet voor wajongers is inspirerend. Ik wens jullie nog veel succes toe met deze en soortgelijke projecten. Mariëlla Muermans, dank voor de gezellige praatjes, maar ook voor alle hulp met sleutels, afspraken, en dergelijke.

Prof. dr. Han Anema, dr. Cécile Boot en Myrthe van Vilsteren, jullie wil ik graag bedanken voor de prettige samenwerking inzake het 'Care for Work'-project waar we ook nog even mee doorgaan. Dank voor jullie vertrouwen in mij als onderzoeker en de snelle acties en reacties wanneer overleg nodig was.

Het Trimbos Instituut. In het bijzonder wil ik Prof. dr. Filip Smit bedanken voor de begeleiding tijdens de analyses van het 'Mental Vitality @ Work'-project. Je hebt me gedurende mijn Trimbos-tijd een enorm welkom gevoel gegeven. Je professionaliteit en hartelijkheid zijn inspirerend. Ik waardeer enorm dat ik, ook nu nog, bij je kan aankloppen. Ik ben er zeker van dat je nog veel jonge onderzoekers zal inspireren! Ook wil ik graag Odile, Joran, Eva, Linda, en alle andere 'Trimbosers' bedanken voor de bijzondere tijd die ik bij jullie heb doorgebracht. Ik heb enorm veel opgestoken tijdens mijn 'verblijf'. Dank hiervoor!

In het kader van het eerder genoemde 'Mental Vitality @ Work'-project wil ik ook graag de dames uit het Coronel Insituut voor Arbeid en Gezondheid bedanken.

Fania en Sara: dank dat we allemaal analyses op jullie data mochten loslaten, en dank ook voor de leuke gesprekken, over allerlei, ze werkten motiverend. Dr. Karen Nieuwenhuijsen en Prof. dr. Judith Sluiter, dankzij jullie oog voor detail verbeterden onze analyses en conceptartikelen.

De Welder groep, en in het bijzonder Catrien, Bruno, Joost, Jenny en Kerst, bedankt voor de samenwerking. Ondanks de spreekwoordelijke beren op onze weg hebben we toch altijd het doorzettingsvermogen weten te behouden. Dankzij het transparante contact hebben we samen mooi wetenschappelijk onderzoek neergezet en een aantal mooie publicaties weten te schrijven.

Kees de Kock, jou wil ik graag bedanken voor je inzet in zowel het 'Huisartsen Werk'- project, de vele taxiritjes naar het station, maar ook voor je expertise en kennis. Jij hebt voorkomen dat ik eindeloos de werking van bepaalde medicijnen moest uitzoeken. Die telefonische afstemming was erg waardevol, en bovendien minder tijdrovend. Dank voor je hulp! Ook jou dr. Romy Steenbeek wil ik hier bedanken voor de energie die je in het 'Huisartsen Werk'-project hebt gestopt, voor de overleggen die je begeleidde en de leuke gesprekken (die dan bijna uitsluitend over paarden gingen). Prof. dr. Toine Lagro-Janssen en Prof. dr. André Knottnerus, hartelijk dank voor jullie input tijdens het schrijven van conceptartikelen en advies inzake de analyses voor het 'Huisartsen Werk'-project. Francine van den Driessen-Mareeuw, mocht ik het in de mails nog niet genoeg benadrukt hebben: bedankt, en alvast bedankt, jouw hulp is onmisbaar.

Voor het nog steeds lopende onderzoek inzake hartrevalidatie wil ik de partners vanuit het Maastricht UMC+ en Adelante Zorggroep bedanken voor hun medewerking aan het onderzoek. Vele handen hebben inderdaad lichter werk gemaakt. Bijzonder veel dank daarom aan Tjerkje, Bianca, Suzanne & Pia (MUMC+) en aan Yvonne, Kim & Jeroen (Adelante). Dr. Petra Kuijpers en Prof. dr. Rob Smeets dank voor jullie input tijdens het schrijven van de onderzoeksvoorstellen en de interesse tijdens het project. We gaan er wat moois van maken.

Care for Companies: Roel Leeuwenburgh, dank voor het vertrouwen en de kansen die je me gegeven hebt. Helaas hebben we vaak nieuwe plannen moeten maken en is er vooralsnog niet meteen uitgerold wat we destijds voor ogen hadden. Toch hoop ik van harte dat we ergens weer een draadje kunnen oppikken en toch onze visies op papier krijgen. Joke, Linda, Sita, Edith, Erwin, Ton en Nienke: we hebben met mekaar erg gedreven geprobeerd het project van de grond te krijgen. Ik hoop dat we te zijner tijd opnieuw de draad kunnen oppakken en dat de meesten van jullie ook weer betrokken zullen zijn. Ik heb onze samenwerking altijd erg constructief ervaren. Jullie maakten graag tijd voor me en ik heb me altijd erg welkom gevoeld in Zoetermeer, dank daarvoor!

DANKWOORD

De vakgroep Sociale Geneeskunde (Maastricht University). Voornamelijk voor de samenwerking met Dr. Inge Houkes en Dr. Nicole Hoefsmits ben ik erg dankbaar. Onze samenwerking reikt verder dan het 'COSS'-project. Inge, bedankt dat je altijd interesse hebt getoond in mijn werk en dat we ook samen aan onderwijs bouwen. Je bent enorm benaderbaar en dat waardeer ik erg. Nicole, jij mag je sinds kort ook doctor noemen. Nogmaals gefeliciteerd! Ik vond het een grote eer je paranimef te zijn. Bedankt, voor je vertrouwen in mij als onderzoeker, als paranimef, als collega, als vriend. Verder wil ik iedereen binnen de vakgroep Sociale Geneeskunde bedanken; bij jullie vakgroep voel ik me altijd welkom. Dank in het bijzonder aan Prof. dr. Hans Bosma en Prof. dr. Frans Feron voor jullie waardering van mijn inzet en het mede mogelijk maken van mijn postdoc. Jullie bijzondere kijk op het onderzoeksleven, waarbij grapjes niet mogen ontbreken, werken erg motiverend. Ook Isel van Noppen wil ik graag bedanken. Lieve Isel, ik weet dat je alleen het dankwoord en de stellingen gaat lezen, dat heb je me al verklapt. Maar net zoals jij een 'open boek' bent is dit proefschrift het ook, lees gerust wat verder, het is best de moeite waard. Dank voor je hulp, het inplannen van afspraken, de vele telefoontjes, maar ook voor je persoonlijkheid. Wat ben jij een vrolijk mens! Houden zo, je weet me altijd op te beuren!

Graag wil ik ook mijn collegae aan de vakgroep Health Services Research bedanken voor hun samenwerking en interesse in mijn onderzoek. Ik wil graag een bedankje richten aan iedereen, van junior tot senior. HSR lijkt wel een grote familie, met hier en daar een verre nicht of neef, maar altijd eentje waarop je kan rekenen.

Een aantal HSR-collegae wil ik graag persoonlijk bedanken. Inge van der Putten, wat heb jij veel werk voor me uit handen weten nemen. Dank voor je inzet! Veel succes ook aan jou met de afronding van je promotietraject. Dr. Ghislaine van Mastrigt, dank voor de samenwerking op vele fronten, maar ook voor de momenten wanneer we niet over het werk praatten. Ik wens je veel succes met je projecten, en vergeet ook zeker niet te genieten van die mooie dochters van je! Dr. Mark Govers, dankzij jouw vertrouwen in mij ben ik een (Colombiaanse) ervaring rijker. Dat zal ik altijd blijven waarderen, net zoals ik de koffietjes altijd weet te waarderen (hoewel we dan ook vaak stiekem wilden dat het koffieapparaat wijn tapte). Brigitte en Suus, merci! Jullie zijn al sinds BEOZ mijn toeverlaat als ik weer eens ergens tegenaan loop, of gewoon, even afleiding zoek. Jullie hebben een gouden hart! Ook Janet en Joanna, dank voor al jullie hulp, en alvast bedankt voor alle hulp die ik nog van jullie ga krijgen. Mitchel van Eeden, mijn BEOZ-kamergenootje. Dank voor zowel je luisterend oor als voor alle grappige momenten die we samen hebben doorgebracht. Ik wens je enorm veel succes met je nieuwe baan. Jij gaat dat ongelooflijk goed doen, ben ik zeker van! De meiden van kamer 0.044! Lieve dames, dank voor al jullie advies en vooral voor alle mooie momenten. Lieve Janneke, even was je toch ook een klein beetje Belg, net zoals ik een klein beetje Nederlander ben. Ik heb veel bewondering voor jouw doorzettingsvermogen en vastberadenheid, en voor

hoe je in het leven staat. Geniet van die twee prachtige kereltjes van je, en van de afronding van je promotietraject. Je gaat dat super doen. Lieve Ariane, bij jou voelt een mens zich toch gewoon goed. Ik heb enorm veel waardering voor jouw positieve kijk op de wereld. Ik hou van die goedlachse wijze waarop jij door het leven gaat. Dat meisje van jullie treft het, jullie worden fantastische ouders! Lieve Reina, of lieve Dr. Reina ;-) Dank ook aan jou voor de leuke 0.044 momenten, en daarbuiten. Met jou in de buurt valt altijd wat te beleven. Ik heb veel bewondering voor de duizendpoot die je bent. Ik heb me meteen 'op mijn gemak' gevoeld bij jou, ook wanneer de taalbarrière weer eens tot hilarische taferelen leidde. Kamer 0.044 dames, laten we ook na Maastricht contact houden! Lieve Joan, met jou klikte het onmiddellijk. Ik hou van die geen-blad-voor-de-mond attitude van jou, gewoon lekker zijn wie je bent. Ik bewonder je gedrevenheid. Jij kan bergen verzetten. Met jou aan mijn zijde komt het vast goed. Dankjewel dat je mijn paranimf wil zijn!

De Academie Medische Vervolgopleidingen (Maastricht UMC+), wat een stel leuke collegae! Dr. Scheltus van Luijk, wij waren meteen een goed duo! Ik heb veel respect en bewondering voor jou, je werk en je kennis. Ik ben er zeker van dat jij me nog veel kan leren. Bedankt dat je me overal bij betreft en dat je ervoor zorgt dat ik mezelf blijf uitdagen. Prof. dr. Laurents Stassen, ook aan jou woorden van dank. Dank voor je vertrouwen in mij. Ondanks je drukke agenda weet ik dat ik altijd bij je terecht kan. Ik waardeer je toegankelijkheid en heb veel bewondering voor hoe jij je taken combineert en mensen weet te inspireren. Ook de dames: Danny, Suzanne, Marie-José, Nicole, Anouk, Yvonne en Marleen, bedankt voor jullie interesse in mijn proefschrift, mijn werk en mijn leven. En bedankt vooral voor jullie bekwaamheid, hulp en adviezen. Ik voel me al helemaal op mijn plek in het ziekenhuis. Bedankt dat jullie me meteen zo een warm en vertrouwd gevoel hebben gegeven! Wat zijn jullie gezellig!

Ook de collegae van KEMTA (Maastricht UMC+) die ik nu veel vaker zie nu we een gang en een koffiezetapparaat delen, bedankt voor jullie interesse in mijn werk. In het bijzonder Dr. Brigitte Essers, bij jou kan ik altijd aankloppen! Ik vind de gesprekken met jou altijd erg inspirerend en motiverend. Prof. dr. Carmen Dirksen en Dr. Manuela Joore, bedankt om jullie expertise te delen en me in deze fase van mijn carrière ook bij te staan met waardevol advies. Ik kijk er alvast naar uit om met KEMTA samen te werken. Dr. Jill Bindels en Dr. Bram Ramaekers, ook jullie zie ik ondertussen weer wat vaker. Het is altijd erg gezellig wanneer we elkaar treffen. Bedankt voor de mooie verhalen die we met elkaar delen.

The CIHR Strategic Training Program In Work Disability Prevention. A special thank you word to Prof. dr. Patrick Loisel who made it possible for me to be part of such an inspiring group of researchers. Dear Sandra Knol, thank you for everything you did to help me arrange my yearly visits to Canada. Your presence and assistance was of tremendous value. I think I can speak on behalf of every WDPI-trainee

DANKWOORD

when I say: 'you are a wonderful person!' Many thanks also to Dr. William Shaw and Dr. Ellen McEachen for acting as my mentors during the optional courses. I learned a lot from working with both of you. Also a huge thank you to the whole group of inspiring trainers and trainees. I feel privileged to know and to have worked with you. A special thanks to Dr. Sheilah Hogg-Johnson, Prof. dr. Emile Tompa, Prof. dr. Carolyn Dewa, Prof. dr. Ute Bultman, Dr. Christian Stahl, Dr. Freek Lotters, Dr. Juliette (Archie) Cooper, Prof. dr. Glenn Pransky, Dr. Katia Costa-Black, Dr. Philippe Mairiaux, you are truly an inspiration.

My WDPI-cohort. What a wonderful, enriching experience that was! Three years in a row: Toronto, with the most amazing group of people I've ever worked with. One by one you guys have amazing personalities. I feel blessed to have met you all, and above all, to be able to call you all my friends! Asa, Emilie, Mette, Ulrik: my Scandinavian friends. Thank you guys so much, for all the wonderful moments we have spent. Genevieve and Susan, my wonderful Australian friends. A very special thank you is in place here, not only for the lovely moments we have spent, the funny things we have did, the great experiences we shared, but also for your help in writing this thesis. Thank you both so much for the English editing of parts of this thesis, your advices and your comments. I'm so looking forward to come and visit you both in Australia! I promise, one day I'll be at your doorstep. Dear Daniella, the sweetest Brazilian girl I have ever met. You have such a warm personality. Do we share some great memories! I know for sure that whenever I want to 'escape' Europe, there will be a place for me to stay in Brazil. Thank you! Sweet 'Dutchies': Hardy en Martine. I feel so fortunate to have met you guys. From the first moment we met, we connected. We share the same interests and never run out of things to talk about. It feels like I know you for ages. Martine, never lose that beautiful spontaneity of yours.

In het bijzonder wil ik ook graag An & Stefan bedanken. Ik weet dat ik altijd op jullie mag rekenen, bedankt voor jullie steun en vertrouwen. Verder ook veel dank aan de 'tuttenclub': Loe & Mariette, Mariet, Ine & Harry, voor de heerlijke etentjes en de gezellige momenten. Merci voor jullie interesse in mijn werk. Ook Martin & Maria, Guido & Rochelio, veel dank!

In dit dankwoord wil ik ook graag mijn vele vrienden bedanken die, soms intens en soms wat meer vanaf de zijlijn, met dit werk hebben meegedeeld. Zonder mijn vrienden zou ik nooit gestaan hebben waar ik nu sta en zou dit vooral een erg saaie reis geweest zijn. Ik wil iedereen daarom uit de grond van mijn hart bedanken voor de interesse maar vooral ook voor de afleiding. Het leven is heel erg de moeite waard met jullie erin! Een spreekwoord zegt: 'Het leven is één groot feest, je moet alleen zelf de slingers ophangen.' Wat kunnen wij goed slingers ophangen!

Lieve Waltwilder-vrienden: de meeste van jullie ken ik via de KLJ, een heerlijke tijd. Nooit zal ik onze hilarische momenten beleven. Ik kijk er enorm naar uit om samen nog meer geschiedenis te schrijven. In het bijzonder: Dennis, Ann, Giel & Elien, David & Lien, Tom & Anneke, Els & Wannes, Elke & Jan, en natuurlijk alle andere KLJ'ers, MERCI!

Lieve Eigenbilzen-vrienden: Chantal & Stefan, Daisy & Robby, Tonia & Matti, Viv'ke & Guido, Mario & Geert, Robert & Diane, Rudi & Alinda, Peter, Johnny & Lizzy, Christel & Frank ook aan jullie 'ne dikke merci'. Bedankt voor alle mooie momenten samen, we gaan er nog meer memorabele maken!

Lieve stallèrkes: Kevin & Sylvia, Kenneth & Dorien, Lexie, Heleen & Kim, Joyce, Henkie, Robrecht & Els, Stijn & Carolina, Yannick, wat zijn we een leuk clubke! Merci voor de fantastische momenten die we al hebben mogen delen, en merci ook voor jullie interesse in mijn werk. Hier is het dus, het is af ☺

Lieve vrienden, jullie zijn stuk voor stuk van onbeschrijfelijke waarde voor mij! Graag wil ik hier de gelegenheid nemen om een aantal van jullie in het bijzonder in het zonnetje te zetten.

Els, wat een mooie cover is het geworden! Ik ben 'super-content', ge zijt een talent! Bedankt, echt!

Anneke & Luc, Hanne & Kristof, wat zijn jullie schatten van vrienden. Met jullie heb ik altijd plezier. Ik zou me serieus geen raad weten moest onze vriendschap ophouden te bestaan. Wat hecht ik veel aan jullie. Jullie hebben het hart op de juiste plaats, en bovendien, dat hart van jullie kleurt ook nog eens goud! Merci lieverds!

Veerle & Kevin, wat ben ik graag bij jullie. De positieve, goedlachse 'vibe' die jullie uitstralen werkt aanstekelijk. Merci voor alle interesse die jullie getoond hebben, de leuke gesprekken en leuke momenten die we delen, en natuurlijk ook voor het 'meterschap' van Coopertje ☺

Joyce en Roberto, een prachtig koppel. Met jullie is het fun verzekerd! Joske, verlies nooit die mooie, spontane, praatgrage persoonlijkheid van jou, ik waardeer 'm enorm!

Nick, mijn maatje! Ik heb veel bewondering voor hoe jij in het leven staat, voor alle stappen die je al hebt gezet en voor alle plannen die je maakt. Ondanks de afstand weet ik dat ik altijd op u kan rekenen, en weet dat dat wederzijds is. Ik wens u echt het allerbeste toe, een leven vol liefde en geluk. In Bilzen, Londen of Australië, you'll always be my best friend!

DANKWOORD

Steven, wat gaan wij ver terug in de tijd. Een onvoorwaardelijke vriendschap, ook al zijn we ongeveer 200 km van mekaar verwijderd. Ik heb veel bewondering voor jou als werkpaard, liefdevolle partner voor Miriam en ongelofelijk tedere vader voor jullie liefste schat Sven. Het is hartverwarmend om te zien wat een warm en hecht gezin jullie zijn. Ik heb me dan ook altijd enorm thuis gevoeld bij jullie in Driebergen. Bedankt ook Miriam dat je me zo nauw weet te betrekken bij de grote stappen die lief Svennetje al zet!

Ellen & Herwig, ook voor onze vriendschap ben ik bijzonder dankbaar! De slingers blijven nog wel een tijdje hangen ;-). Liefste Ellen, ik kijk enorm uit naar jullie huwelijk en voel me erg vereerd dat ik je getuige mag zijn. Bedankt voor alle mooie momenten, en vooral ook voor de liefde en vriendschap die ik van je krijg!

Sara & Kim, Harold & Paula, jullie maken dit mede mogelijk. Misschien zonder het goed en wel te beseffen, maar zonder jullie hulp zou ik never-te-nooit op mijn gemak zijn wanneer ik ergens hoog in Nederland of ver overzee ben. Dank dat ik altijd op jullie mag vertrouwen en op jullie mag rekenen. Daarom, ook in naam van mijn allerliefste schat Astar: 'ne welgemeende dikke merci!

Lieve Carmen, lieve Katrien, lieve Chloé, mijn rotsen in de branding, my girls. Wat ben ik dankbaar voor jullie in mijn leven. Laten we voor altijd vriendinnen blijven!

Lieve Carmen, onze humor reikt verder dan de meesten kunnen begrijpen. Je bent een heerlijke vriendin: goedlachs, met een luisterend oor en altijd in-for-a-party. De herinneringen die wij delen, de momenten die wij beleefd hebben, onze schaterbuien, onze fratsen, wat een heerlijk leven is het met jou erin! We spreken één ding af: laten we nooit volwassen worden!

Lieve Katrien, we kennen elkaar al van toen we kleine pagadderkes waren, ganse dagen gependend op de manege. Het verrijkt mijn leven te weten dat ik altijd, no-matter-what, op jou kan rekenen. Ik heb enorm veel bewondering voor je wilskracht en de manier waarop je in het leven staat. Je weet wat je wil, en daar ga je voor. Kaatje, ik hecht immens veel waarde aan onze vriendschap, en onze eeuwige gesprekken over honden en paarden ☺ wat delen wij bijzondere momenten samen. Stelling 10 is voor ons!

Liefste Chloé, wat gaan ook wij ver terug in de tijd. En wat hebben wij al veel avonturen beleefd! Ook al laat je buitenkant dat niet altijd uitschijnen, je hebt echt een klein, gouden hartje. Je staat altijd voor me klaar en nooit is je wat te veel. Je bent mijn reddertje in nood, mijn afleiding, mijn rots! Het geeft me veel vertrouwen te weten dat je naast me staat tijdens mijn verdediging: bedankt dat je altijd in me bent blijven geloven! Gij en ik tegen de wereld!

Graag wil ik ook in dit dankwoord mijn schoonfamilie bedanken. Dankjewel tant Paulette, bompa, Elviera, Kirsten, en de liefste tweeling van de wereld: Ties & Stien, voor alle mooie momenten die we de afgelopen jaren al hebben beleefd. Vooral ook heel veel dank voor jullie interesse en medeleven in dit werk. Het is af ☺ Lieve Ties & Stien, ik ben blij dat ik jullie “tante’ke” ben mogen worden. Ik geniet ervan jullie te zien opgroeien!

Liefste schoonouders, Teejoo & Viviane: onbeschrijflijk veel dank voor alles wat jullie voor mij, voor ons, doen. Ik waardeer enorm dat jullie altijd voor ons klaar staan en dat jullie mij een tweede thuis hebben gegeven! Met jullie kan ik over alles praten en vooral ook heel veel plezier maken. Ik prijs me enorm gelukkig dat ik zulke lieve, joviale en sociale schoonouders heb mogen treffen. Ik hou van jullie!

Lieve familie “papa’s kant”: tant Miet, nonk Gerard, Johnny, Vicky, Lucy & Amy, Jan, Stephanie, Kristel, Kenny, en lief vechtertje Lobke. Het geeft me veel vertrouwen te weten dat ik ook altijd op jullie mag rekenen. Bedankt voor jullie interesse in mijn werk en voor de fijne momenten die we delen. Lieve Kristel en Kenny, ik heb veel respect voor de manier waarop jullie zijn blijven doorzetten. Jullie zijn fantastische ouders, Lobke had niet beter kunnen treffen! Ik ben trots op jullie.

Familie “mama’s kant”: Lieve nonk Eric, tant Monique en Bart, bij jullie voel ik me altijd welkom. Bedankt voor de hechte band die we hebben en de leuke tijden die we samen beleven. Lieve Ingrid, Fons, Steven & Yana en Koen, dank jullie wel voor alle grappige, leuke en bijzondere momenten die wij samen gespendeerd hebben. Lieve Steven en Koen, jullie zijn mijn favoriete neefjes, dat weten jullie toch hé (niet tegen Bart zeggen hoor ☺). Tata’ke, wat een bijzonder lieve tante jij bent. Jij staat altijd voor me klaar en ik weet dat je enorm trots op me bent. Dankjewel voor alle liefde die je me geeft! Bonke en Bonnemie, mijn grootouders. Kijk, dit is nu waar ik zo lang aan heb gewerkt. Ik weet dat jullie ook zonder dit boekje heel erg trots op me zijn en dat doet me goed! In het bijzonder wil ik bonnemie bedanken, mijn meter. Ook uwe naam pronkt op dit boek! Ge zijt een sterke vrouw waaraan ik weet wat ik heb. Lieve bonnemie, ik hoop uit de grond van mijn hart dat ge nooit zult opgeven en dat ge altijd zult proberen aan mijn zijde te staan. Ik wil nog heel wat mijlpalen zetten en kan me niet voorstellen dat gij daar niet bij zou zijn. Met u in de buurt is het nooit saai, wat moet ik altijd lachen met uw ‘fratsen’. Lieve bonnemie, lieve familie, ik hou van jullie, onvoorwaardelijk veel!

De volgende dankwoorden richt ik graag aan mijn ouders en broertje. Jullie zijn van onschatbare waarde voor mij. Jullie geven mijn leven kleur en betekenis, en ik kan me geen betere ouders voorstellen! Ik ben enorm dankbaar dat ondanks het feit dat jullie, mama en papa, niet meer bij mekaar zijn, wel altijd als één front achter me zijn blijven staan. Vele ouders zouden aan jullie een voorbeeld mogen nemen!

DANKWOORD

Lieve papa, jij bent de lijm die ons samenhoudt, onze steun en toeverlaat. Ik heb veel respect voor u papa. Voor de manier waarop je altijd positief probeert te blijven, hoe hard het leven soms ook voor je is. Bedankt dat je altijd voor me klaar staat. Ik waardeer enorm dat je me altijd goede raad weet te geven en dat je altijd het beste met me voor hebt, ook al wil ik dat soms niet meteen inzien. Papa, ik hou van u! Lieve Brigitte, ook voor u alleen maar liefde. De band die wij hebben is erg bijzonder, en vooral onbreekbaar. Je bent een tweede mama voor me. Ik heb enorm veel respect voor uw doorzettingsvermogen en wilskracht. Je bent een sterke en dappere vrouw en ik kijk enorm naar je op. Het vult mijn hart te zien hoe goed je bent voor mij, voor papa en voor Niek. Dankjewel lieve Brigitte! Niek, lieve Niek, mijn kleine broertje. Wat gaat de tijd snel! 15 word je al bijna. Ik herinner me nog zo goed toen ik je voor de eerste keer in mijn armen hield. Ik ben een trotse zus en een trotse meter! Je bent een lieve kerel, met het hart op de juiste plaats! Ik hoop dat je weet dat ik altijd voor je klaar zal staan en dat je altijd bij me terecht kan. Je blijft toch mijn 'kleine broertje'. Ik hou van je!

Lieve mamie, mijn grote voorbeeld, mijn eeuwige steun en toeverlaat, dit boek draag ik op aan jou. Jij bent altijd in mij blijven geloven, ook wanneer ik dat zelf niet meer deed. Bedankt voor je onvoorwaardelijke liefde. Ik ben erg dankbaar dat je me vrij hebt gelaten om mijn eigen weg te zoeken, maar vooral ook dat je me altijd bent blijven steunen, zelfs wanneer ik de weg even kwijt was. Mamie, ik weet dat je het soms moeilijk had wanneer ik weer eens voor een tijdje naar het buitenland ging, maar kijk, het is helemaal goed gekomen. Veel dank dat je me alle kansen hebt laten grijpen, ook al moest je daar zelf soms wat voor laten. Mocht ik later nog maar half zo een goede mama worden dan jij bent, dan zal ik nog de beste zijn. Ik kan me echt geen betere mama voorstellen. Je bent onmisbaar in mijn leven! Wij blijven vriendjes, jij en ik, tot de laatste snik! Ik hou van u!

Liefje, jou bewaarde ik tot de laatste. Ik zou inderdaad zo opnieuw 300 pagina's kunnen vullen met hoeveel jij voor me betekent ;-). Ik bewonder je spontane karakter en vind het heerlijk hoe jij me laat lachen. Bedankt dat je me andere perspectieven hebt leren waarderen en me soms uit mijn 'comfort-zone' haalt. Jij maakt van mij een beter mens. De kaft van dit boek, strak en in hooglans wit, heb ik voor jou uitgekozen (misschien toch een goed idee voor onze keuken later). Het is bijzonder hoe wij mekaar aanvullen en het geeft me een heerlijk gevoel te weten dat je me graag ziet. Sinds jij in mijn leven bent is het compleet. Lieve Wim, onze reis is nog lang niet ten einde, je bent de man van mijn leven, ik wil voor altijd bij je blijven. Ik hou van je, tot aan de maan en terug!

Bilzen, 8 mei 2015

Curriculum Vitae

Cindy Noben was born in Tongeren, Belgium, on June 1st, 1986. After completing her secondary education she studied General Health Sciences at Maastricht University for which she obtained her bachelor degree in 2009. During the bachelor (2008) she was enrolled in the Erasmus Exchange Program to participate in Advanced Courses for Research in Public Health at Mälardalen University, Västerås, Sweden.

In September 2009 Cindy enrolled in the master program Health Policy, Economics and Management at Maastricht University. During the master she conducted field research at Hospital Universitario Mayor Méderi, Bogotá, Colombia and combined this work with finishing her master thesis at Universidad del Rosario, Bogotá, Colombia.

After obtaining her master's degree in August 2010, she started working as a PhD candidate at the research school CAPHRI, department Health Services Research of Maastricht University. Cindy enrolled in the large research project entitled 'Arbeid & Zorg' funded by the Dutch Foundation Institute Gak. During her PhD trajectory she got involved in various research projects and working groups. From 2012 to 2014, she got granted to enroll in the CIHR Strategic Training Program in Work Disability Prevention at the University of Toronto, Dalla Lana School of Public Health, Canada. Cindy was also appointed as an external researcher at the Netherlands Institute of Mental Health and Addiction (the Trimbos Institute) from September 2012 until May 2013. Aside from her research activities she performed various educational roles such as tutor, trainer and lecturer within the bachelor and master programs in Health Sciences offered at Maastricht's Faculty of Health, Medicine and Life Sciences.

Since October 2014, Cindy works in a post-doc function at CAPHRI within the research programs 'Health Technology Assessment' and 'Social Participation and Health' of Maastricht University. She is currently involved as a module planning group member of the course on Health Policy General Principles Applied to Local Settings offered within the bachelor program General Health Sciences, and as a planning group member of the Master program Work, Health and Career. Cindy currently also functions as region coordinator (South-East Netherlands) at the Maastricht School of Post Graduate Medical Education of the Maastricht University Medical Centre.

Scientific contributions, grants & awards

Scientific journal contributions

Publications

- **Cindy Noben**, Frans Nijhuis, Angelique de Rijk, Silvia Evers. *Design of a trial-based economic evaluation on the cost-effectiveness of employability interventions among work disabled employees or employees at risk of work disability: The CASE-study*. BMC Public Health. 2012; 12(1):43
- **Cindy Noben**, Filip Smit, Karen Nieuwenhuijsen, Sarah Ketelaar, Fania Gärtner, Brigitte Boon, Judith Sluiter, Silvia Evers. *Comparative cost-effectiveness of two interventions for improving work functioning among nurses: pragmatic cluster randomised trial*. International Journal of Nursing Studies. 2014, 51:10.
- **Cindy Noben**, Silvia Evers, Frans Nijhuis, Angelique de Rijk. *Quality appraisal of generic self-reported instruments measuring health-related productivity changes: A systematic review*. BMC Public Health. 2014, 14:115.
- Ben Wijnen, Inge van der Putten, Siebren Groothuis, Reina de Kinderen, **Cindy Noben**, Aggie Paulus, Bram Ramaekers, Gaston Vogel, Michael Hiligsmann. *Discrete-Choice Experiment versus Rating Scale Exercise to Evaluate the Relative Importance of Attributes*. Expert Review of Pharmacoeconomics & Outcomes Research. 2015;1-8.
- **Cindy Noben**, Filip Smit, Karen Nieuwenhuijsen, Sarah Ketelaar, Fania Gärtner, Judith Sluiter, Silvia Evers. *Protecting and promoting mental health of nurses in the hospital setting: Is it cost-effective from an employer's perspective?* Accepted for publication by International Journal of Occupational Medicine and Environmental Health

Revised submissions

- **Cindy Noben**, Angelique de Rijk, Frans Nijhuis, Jan Kottner, Silvia Evers. *Patient self-reported and administrative data used for cost estimations of healthcare resource use: A systematic review of methodological quality and validation evidence*. Journal of Clinical Epidemiology (first revision)

SCIENTIFIC CONTRIBUTIONS, GRANTS & AWARDS

- **Cindy Noben**, Silvia Evers, Frans Nijhuis, Joost van Genabeek, Angelique de Rijk. *A pilot economic evaluation to improve a web-based employability intervention for employees*. Journal of Occupational Health (first revision)
- **Cindy Noben**, Nicole Hoefsmit, Silvia Evers, Angelique de Rijk, Inge Houkes, Frans Nijhuis. *Economic evaluation of a new organizational RTW intervention to improve Cooperation between Sick-listed employees and their Supervisors (COSS): A field study*. Journal of Occupational and Environmental Medicine (first revision)

Submissions

- Nicole Hoefsmit, Inge Houkes, Nicolle Boumans, **Cindy Noben**, Bjorn Winkens, Frans Nijhuis. *The effectiveness of an intervention to enhance cooperation between sick-listed employees and their supervisors (COSS)*. Journal of Occupational Rehabilitation.

Other publications

Reports

- In Dutch: *Eindrapportage Wajongproject "Een passende baan bij het Slotervaartziekenhuis"*. Maastricht University, 2014. Gemma van Ruitenbeek, Marike Mulder, Philippe Nelissen, Ghislaine van Mastrigt, **Cindy Noben**, Silvia Evers, Fred Zijlstra & Frans Nijhuis.
- In Dutch: *Biedt inzicht uitzicht? Verslag van een onderzoek naar het gebruik van de interventie 'Wijzer met Welder' door een groep leden van CNV Publieke Zaak*. TNO Hoofddorp, 2014. Joost van Genabeek, **Cindy Noben**, Bram Rooijackers, Maartje Bakhuys Roozeboom, Jenny Huijs, Karen Oud Hengel, Anneke Huson, Agnes Meershoek & Angelique de Rijk.
- In Dutch: *Onderzoeksprogramma Arbeid & Gezondheidszorg: Inspiratie voor werkconferentie 'geïntegreerde arbeidsgerelateerde zorg'*. KPMG Plexus, 2015. David Ikkersheim, Frans Nijhuis, Romy Steenbeek, Femke van der Voort, Nicoe Hoefsmit, **Cindy Noben**, Kees de Kock & Myrthe van Vilsteren.

Daily journal

- In Dutch: *"Wajonger bespaart bedrijf geld"*. De Volkskrant, October 29th 2014. Based on report "Een passende baan bij het Slotervaartziekenhuis".

Newsletter webpage

- In Dutch: *"Aandacht voor psychische gezondheid werknemer loont"*. Trimbos nieuws, May 2nd 2013. Based on *"Comparative cost-effectiveness of two interventions for improving work functioning among nurses: pragmatic cluster randomised trial"*.

Scientific conference contributions

Organized sessions

- **Cindy Noben**, Bram Ramaekers, Joanna Thorn, Silvia Evers, William Hollingworth. *Methodological Challenges in Healthcare Resource Use Measurement: Towards Developing Good Practice*. Topical Seminar at the international Health Economics Association/European Conference of Health Economics, 10th World Congress. July 14 – 16, 2014, Dublin, Ireland.
- **Cindy Noben**, Silvia Evers, Angelique de Rijk, Frans Nijhuis. *Employees' responsibilities in work disability prevention – could they?* (2014). Topical Seminar at the Third Scientific Conference on Work Disability Prevention and Integration. September 29 – October 1, 2014, Toronto, Canada.

Presentations

- **Cindy Noben**. Werkhervatting bij hartpatiënten. Invited speaker at 'Continuous Nursing Education' door de Nederlandse Vereniging voor Hart- en Vaat Verpleegkundigen (NVHVV). 21 april 2015. Utrecht, Nederland.
- **Cindy Noben**, Angelique de Rijk, Frans Nijhuis, Silvia Evers. *Self-report methods to assess costs: how valid are they?* (2012) Gaceta Sanitaria vol. 26 – Especial Congreso 2 – Junio 2012, Bilbao, Spain.
- **Cindy Noben**, Silvia Evers, Angelique de Rijk, Frans Nijhuis. *Systematic review of generic self-reported health-related workplace productivity loss instruments (2012)* Gaceta Sanitaria vol. 26 – Especial Congreso 2 – Junio 2012, Bilbao, Spain.
- **Cindy Noben**, Frans Nijhuis, Angelique de Rijk, Silvia Evers. *Design of a trial-based economic evaluation on the cost-effectiveness of employability interventions among work disabled employees or employees at risk of work disability: The CASE-study (2012)*. Gaceta Sanitaria vol. 26 – Especial Congreso 2 – Junio 2012, Bilbao, Spain.
- **Cindy Noben**, Silvia Evers, Frans Nijhuis, Angelique de Rijk. *Quality appraisal of generic self-reported instruments measuring health-related productivity changes: A systematic review*. Progress in health economics, 9th European conference on health economics (ECHE), July 2012, Zürich, Switzerland.

Poster presentations

- The CASE-study: cost effectiveness sustainable employability (Maastricht University, CAPHRI research meeting, the Netherlands. September 2011)
- Cost effectiveness of return to work interventions (RTWi): A research protocol (Maastricht University, CAPHRI research meeting, the Netherlands. April 2012)

SCIENTIFIC CONTRIBUTIONS, GRANTS & AWARDS

- Comparative cost effectiveness of two interventions for improving work functioning among nurses: pragmatic cluster randomized trial (Xth ENMESH international conference mental health care in Europe, Verona, Italy. October 3-5, 2013)

Grant

- Honored trainee at the “*Work Disability Prevention CIHR Strategic Training Program*”, University of Toronto, Dalla Lana School of Public Health, Canada

Award

- Award for Best Poster at the Xth ENMESH international conference mental health care in Europe, Verona, Italy. October 5th 2013. “*Comparative cost-effectiveness of two interventions for improving work functioning among nurses: pragmatic cluster randomised trial*”

In de Nederlandse participatiesamenleving wordt gestreefd naar duurzame inzetbaarheid van werknemers, ook van diegenen met gezondheidsproblemen. Er worden verschillende interventies ontwikkeld om de duurzame inzetbaarheid van deze groep te vergroten. Voor werkgevers en beleidsmakers is het van belang inzicht te hebben in de kosteneffectiviteit van deze inzetbaarheidsinterventies alvorens ze te implementeren. Kosteneffectief betekent dat netto-opbrengsten in een gunstige verhouding staan tot de netto-kosten. Er is echter geen consensus over de methoden om dit te onderzoeken. Dit proefschrift levert een bijdrage aan de huidige wetenschappelijke kennis ten aanzien van het ontwikkelen, evalueren en implementeren van inzetbaarheidsinterventies. Methodologische overwegingen, alsook de kosteneffectiviteit van bestaande en innovatieve inzetbaarheidsinterventies worden beschreven.



CINDY NOBEN (1986) is werkzaam als postdoc onderzoeker en als docent aan de Faculty of Health, Medicine and Life Sciences aan de Universiteit Maastricht. Ze is tevens werkzaam als regionaal projectcoördinator 'doelmatigheid van zorg' in het Maastricht UMC+. Als Erasmus Exchange student studeerde Cindy aan Mälardalen University, Västerås, Zweden. Tijdens haar masteropleiding werkte ze in Hospital Universitario Mayor en schreef ze haar thesis aan Universidad del Rosario, Bogotá, Colombia. Gedurende haar PhD-traject was ze research fellow binnen het CIHR Strategic Training Program in Work Disability Prevention aan de University of Toronto, Canada. Gedurende haar promotietraject voerde Cindy voor negen verschillende projecten economische evaluaties uit waarvan er in dit proefschrift vijf gepresenteerd worden.