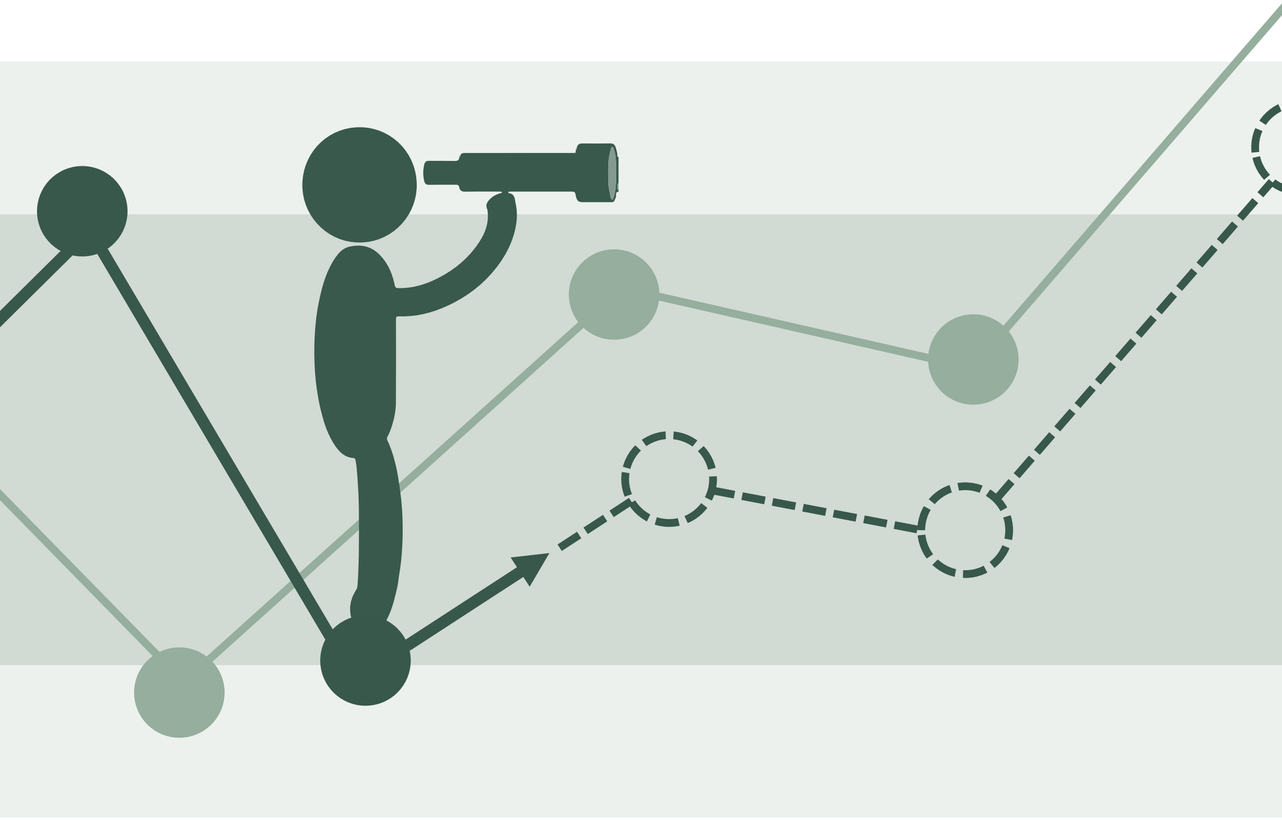
PREDICTIVE MODELLING IN INSURANCE MEDICINE



Ilse Louwerse

Predictive modelling in insurance medicine

Ilse Louwerse

ISBN: 978-94-6416-055-0
Layout and cover design: Ilse Louwerse
Printing: Ridderprint

The studies in this thesis were carried out at the Department of Occupational and Public Health of Amsterdam UMC, location VU University Medical Center (VUmc), and the Amsterdam Public Health research institute, the Netherlands. The Department of Occupational and Public Health participates in the Dutch Research Center for Insurance Medicine, which is a joint initiative of the VUmc, the Dutch Institute for Employment Benefit Schemes (UWV), the Amsterdam Medical Center, and the University Medical Center Groningen. The studies in this thesis were funded by UWV. Financial support for the printing of this thesis has been kindly provided by UWV.

VRIJE UNIVERSITEIT

Predictive modelling in insurance medicine

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op dinsdag 17 november om 11:45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Ilse Louwerse
geboren te Rotterdam

promotoren:	prof.dr. J.R. Anema
	prof.dr. A.J. van der Beek
copromotoren:	dr. M.A. Huijsmans
	dr. H.J. van Rijssen

Table of contents

CHAPTER 1

General introduction

"I think that data analytics and prediction models should increasingly be put into practice in insurance medicine."

"I could use a prediction model during work disability assessments to verify whether the prognosis that I make based on literature and guidelines is correct."

"To be useful in practice, it is important to know exactly how the prediction model has been developed, how each of the prognostic factors contributes to the prediction for a single claimant."

"Given the limited occupational health resources, could data analytics guide insurance physicians when they need to decide on the optimal allocation of the available time?"

These are quotes from managers and insurance physicians (IPs) of the Dutch Social Security Institute (SSI). The quotes illustrate the opportunities of using data analytics and prediction models in insurance medicine in the Netherlands. However, also some of the considerations that IPs have and barriers that need to be overcome for successful implementation in practice are mentioned. Despite the increasing power and potential benefits of big data and prediction models, they are currently not systematically used in disability assessments in insurance medicine in the Netherlands. Therefore, the overall aim of this thesis was to get insight into the opportunities and challenges of data analytics in work disability assessments and rehabilitation support.

Social security and insurance medicine in the Netherlands

In most European countries, workers who are sick listed can rely on financial support to compensate loss of income and on support in returning to work [1]. For sick listed workers in the Netherlands, this process is illustrated in the top part of figure 1. During the first two years of sickness absence, employers are responsible for continued payment of wages and rehabilitation support for employees with an employment contract. In contrast, workers without an employment contract (unemployed and temporary agency workers) can apply for a sickness absence benefit at the SSI under the Sickness Benefits Act [2]. In 2019, about 298,000 individuals were granted a sickness absence benefit by the SSI [3]. Sickness absence benefits equal 70% of the last daily wage. They last as long as claimants are sick listed, with a maximum duration of two years. As sick listed workers without an employment contract do not have a workplace to return to, outflow from the sickness absence benefit does often not result in actual return to work [4].

After two years of sickness absence, both workers with and without an employment contract can apply for a work disability benefit under the Dutch Work and Income Act (WIA). This is shown in the bottom part of figure 1. As is the case for sickness absence benefits, work disability

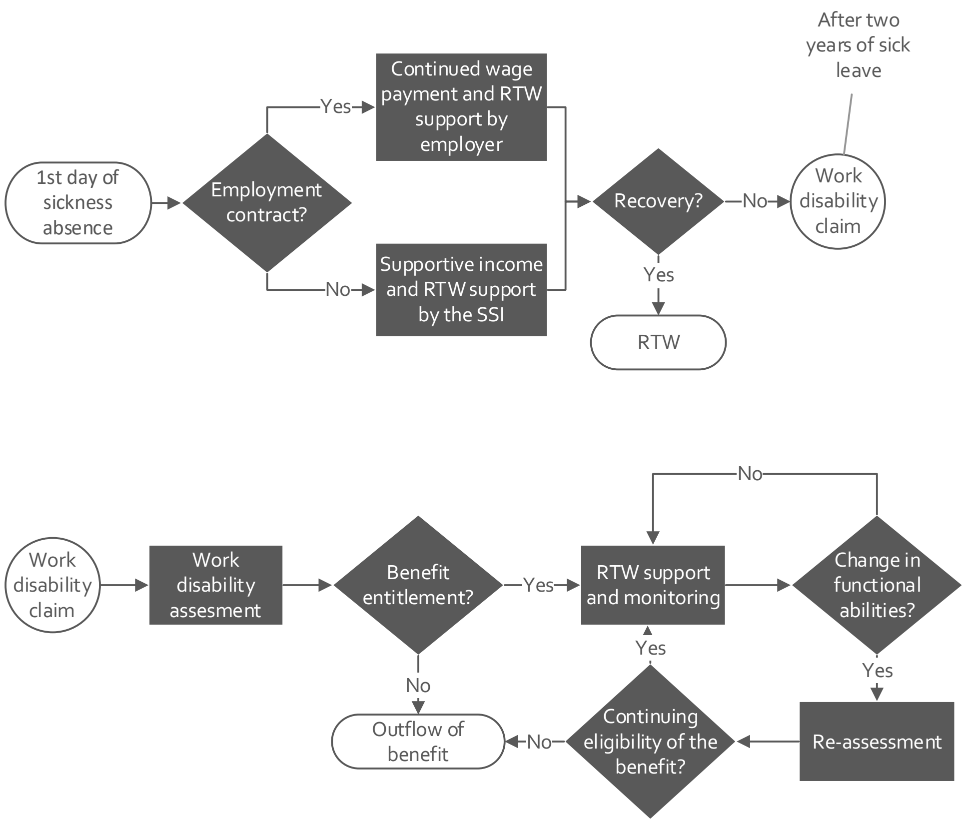


Figure 1. Flow chart of the sickness absence and work disability process

benefits can be approved for a disease or handicap due to either non-occupational or occupational causes. If a claimant applies for a work disability benefit, a medical assessment of functional abilities is conducted by an IP, who is employed by the SSI. Depending on the functional abilities listed in the IP's report, there may also be an assessment by a labor expert (LE) who calculates the loss of former wages. A work disability benefit is granted if loss of income exceeds 35% of former wages. About 46,000 individuals were granted a work disability benefit by the SSI in 2019 [3]. Eligibility for work disability benefit continues either until the SSI receives a notification that someone has recovered, or when an IP assesses substantial improvement of the medical situation such that functional abilities improve.

In the Netherlands, as in many other countries, IPs perform a crucial role in the sickness absence and work disability process. Based on IPs' assessment of diagnoses and functional limitations, it is determined whether a benefit should be granted or not. During the medical disability assessment, IPs also need to estimate prognosis of future changes in work disability as, once a work disability benefit has been granted, changes in health may alter its continuing eligibility [1, 5]. IPs conduct medical re-assessments to determine whether a claimant's health including functional abilities has improved or deteriorated to such an extent that adjustment of the benefit or support to return to work is required. In the Netherlands, IPs need to determine during the medical work disability assessment if and when a re-assessment should be planned. In general, IPs consider this as the most difficult part of the work disability assessment as it requires rather complex predictions, in which a broad range of factors play a role [6, 7]. However, accurate prognosis of future changes in work (dis)ability is important to identify those in need for return to work interventions and to ensure that medical re-assessments are planned at the time an interview with an IP or LE has most added value.

Facts and figures of work disability

Work disability occurs when someone is unable to perform occupational tasks due to a disease or disorder. It is defined as the gap between personal capabilities and the roles and tasks expected of an individual in the working environment [8, 9]. The dimension and severity of work disability is not only determined by the physical and mental limitations that someone experiences, but by several other factors as well. These include, for instance, a person's experience of the situation, the reactions and expectations of others (especially family members, friends, and colleagues), and the possibilities to adjust the working environment and working arrangements [10]. This means that not all health-related limitations lead to work disability and that individuals that suffer from the same disease or disorder might experience different levels of work disability.

In some cases, work disability can be solved with workplace adjustments, such as allowing flexible working arrangements, providing a ramp, or purchasing speech recognition software. However, in many cases such adjustments are not possible or insufficient for job retention. In that case, individuals can apply for a work disability benefit. As benefits usually only partially cover the loss of income that someone experienced, and individuals with disabilities have extra costs directly related to the presence of the disability, work disability often leads to loss of spendable income. Moreover, once a work disability benefit has been approved, the probability of returning to work is low [11]. Long-term unemployment or occupational inactivity can be bad for an individual's health, especially for those with mental health conditions, and returning to work is generally associated with health improvement [12-14]. Therefore, early identification of specific at-risk groups and implementation of effective interventions to prevent long-term occupational inactivity is important.

Today, work disability is one of the greatest societal and labor market challenges for policy makers in most OECD countries. Public spending on work disability benefits has become a serious financial burden. On average, about 6% of the working-age population relies on disability benefits and OECD countries spend on average 1.2% of their GDP on work disability benefits [11]. This number rises to 2% when sickness absence benefits are included as well. In some countries, e.g. the Netherlands, Norway, Denmark and Sweden, expenditures are even higher, ranging from 3.1 to 4.8% of GDP. Although long-term sickness absence makes up only a small proportion of absences, it accounts for more than one-third of days off and 75% of sickness absence costs [15]. Hence, prevention of work disability and support for returning to work are in the interest of individuals and society as a whole.

Work disability benefit rates have risen in the past decades and are expected to rise even further in the future. One of the main reasons for this rise is the aging population, resulting in a sharp increase in the proportion of the working population who are over the age of 60 years [16]. To improve sustainability of pension schemes, many countries have started to implement statutory retirement reforms that aim at increasing labor force participation of older workers. However, the prevalence of disabilities and medical conditions that leave individuals unable to work generally increase with age. If they work longer, workers have an increased risk of work disability. Work disability benefits may become a substitute for early retirement for older workers who are unable to work until the increased statutory retirement age [17]. Moreover, the probability to return to work decreases with age, and the total number of sickness absence days increases [18]. As both the incidence and prevalence of work disability are higher for older workers, a further increase in work disability benefit rates is expected in the near future.

These developments will pose an even larger burden on the already limited capacity of the occupational health system. The number of IPs working at the SSI as well as the number of doctors that want to follow the postgraduate education in insurance medicine is limited [19]. To be able to serve all benefit claimants now and in the future, there is a high demand for IPs. With the exodus of the baby boomers from the work force, resulting in the retirement of many IPs in the next years, the shortage of IPs will increase even further [20]. This poses additional challenges to the occupational health system and highlights the need for solutions to make the system more effective, e.g. by delegating tasks from IPs to practice staff, using new technologies or predictive analytics.

Gaining further insights into factors associated with long-term sickness absence and work disability might be one of these solutions. In this context, work ability is an important concept. It is defined as the physical, mental and social fit of an individual with the work demands and capability to participate in work [21]. Work ability is commonly measured with the Work Ability Index (WAI), a questionnaire on the demands of work, an individual’s health status and resources [21, 22]. The application of the WAI in research as well as in occupational health services has generally been recognized [23, 24]. Self-assessed work ability is a strong predictor for sickness absence duration and return to work [25, 26]. The concept might help occupational health professionals to better evaluate what is needed in order for sick listed individuals for (partial) return to work, to determine the aspects that could be changed or improved, and to compose a rehabilitation program [27]. As work ability does not necessarily relate an individual's work capacity to the present job, but can consider any kind of jobs found on the labor market, the concept is also relevant for workers without an employment contract [28].

Decision support tools in occupational and insurance medicine

Early identification of individuals at risk of long-term sickness absence and work disability, and an overview of factors associated with sickness absence duration, can help occupational health professionals to target specific at-risk groups and identify effective interventions. For effective planning of re-assessments and allocation of limited resources to support return to work, it may be helpful to quantify which individuals are likely to benefit most from it. For instance, which individuals have an increased risk for long-term sickness absence? Which claimants with a work disability benefit are most likely to experience future improvement in work ability as proxy measure for (increase in) RTW? Is it possible to allocate sick listed workers into distinct groups and offer return to work interventions tailored at the characteristics of these groups?

Prediction models may help to answer the above questions as they enable objective and standardized quantification of a claimant's probability to experience a certain outcome in the future (e.g. long-term sickness absence or work disability). They are useful for the identification of specific groups at risk, and may assist professionals in standardizing the decision making process. Throughout the last decade, several prediction models have been developed focusing on sick-listed workers with an employment contract [29-31]. However, prediction research is relatively new to the field of insurance medicine.

The iterative process of building a prediction model is shown in figure 2. First, the study objective and requirements need to be defined. Based on these, it has to be determined which data is needed and which data sources can be used. The data preparation phase covers all activities from collecting the initial data to cleaning the data and constructing the final dataset. Data can be collected from different sources, for instance registration data or self-reported questionnaires. Before we can build a model, we should transform the data so that it meets the specific requirements of the modelling techniques that we want to apply. This covers, for instance, dealing with missing data, handling categorical data and scaling. Next, appropriate modelling techniques can be applied. As it is usually unknown beforehand which is the best modelling technique, several models will be created and tested. These models can be used to generate predictions, for instance by classifying subjects into different categories based on their

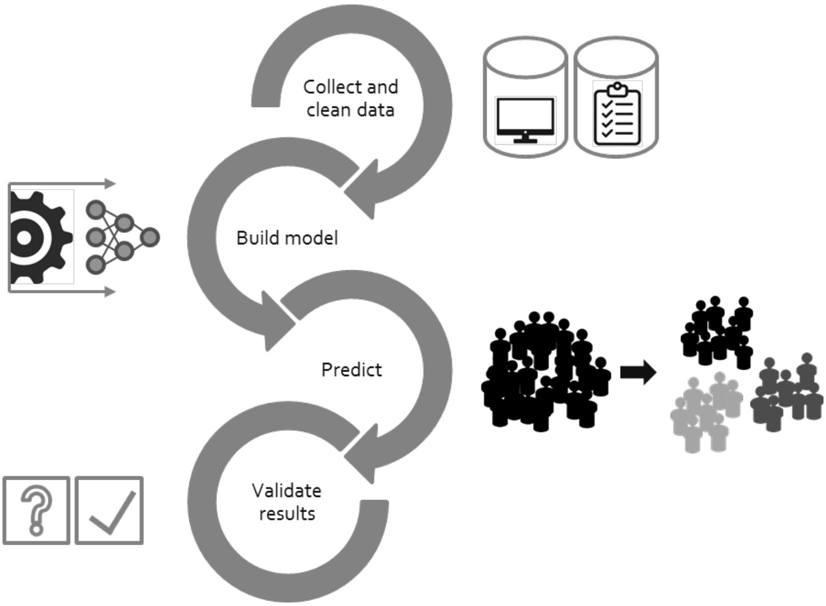


Figure 2. Steps used in building a prediction model

characteristics. The final step is to validate the model predictions to ensure that they generalize against unseen data and that the models meet the practical requirements, and to decide about model deployment. As new insights emerge and knowledge is gathered, throughout the process there can be a loop back to one of the previous steps.

Once a prediction model has been selected for implementation, clear presentation of the outcome of the model is important to ensure proper use in practice. Prediction models are usually presented to professionals as part of a clinical decision support tools (suitable interfaces that display the outcome of a prediction model). They match the characteristics of individual patients with clinical knowledge base or a decision rule, and present patient-specific assessments or recommendations to clinicians. Clinical decision support tools make the results of prediction models easily accessible and interpretable for professionals [32]. These tools are designed to support decision-making; they can introduce efficiencies into the system, optimize the time with the client, improve the overall quality of services and return to work interventions, and increase occupational health quality and efficiency [33-35]. Although such tools are more common in clinical practice, they are currently lacking in insurance medicine in the Netherlands. To be of added value in practice, not only the evidence base underlying these tools needs to be relevant and of high quality, but it is also important that the tool itself is easily accessible and interpretable. Therefore, anticipating on the needs of IPs and other professionals regarding the preferred way of use and design are key components when developing an implementable and effective decision support tool [36].

Aim of this thesis

The main aim of this thesis is to investigate how data-driven insights, prediction models and decision support tools can help IPs in the Netherlands in making evidence-based decisions regarding prognosis of work ability and support to return to work. The primary objectives of this thesis are:

1. To give an overview of factors associated with work disability entitlement and duration;
2. To predict risk of long-term sickness absence and identify distinct subgroups of sick-listed workers without an employment contract;
3. To develop a prediction model and decision support tool predicting future changes in work ability of work disability claimants;
4. To get insight into the efficacy of the decision support tool and IPs' attitudes towards use of the tool.

The answers to these questions may assist IPs and other occupational health professionals to make more accurate decisions regarding prognosis of claimants with a sickness absence and work disability benefit. Moreover, the answers may help to target groups of individuals who are at risk for long-term sickness absence and work disability, identify effective return to work interventions for these groups and increase efficiency.

Outline of this thesis

The thesis is organized as follows:

- Chapter 2 provides an overview of the sociodemographic characteristics and diagnoses of individuals who have been granted a work disability benefit, and examines the duration of their benefits.
- The results of a longitudinal study on prognostic factors for long-term sickness absence among sick listed workers without an employment contract are presented in chapter 3. Moreover, this study describes factors that could be used to identify distinct subgroups of sick listed workers.
- A prediction model for changes in work ability of claimants one year after approval of the work disability benefit is presented in chapter 4.
- Chapter 5 describes the results of a focus group study to explore the preferable way of use and design of a decision support tool for IPs and LEs, based on the prediction model presented in Chapter 4.

- Chapter 6 describes whether use of such a tool affects IPs' prognosis of work ability and their prognostic confidence, and assesses IPs' attitudes towards use of the tool in practice.
- Chapter 7 summarizes and discusses the main research findings, and considers the implications for research, policy and practice.

References

1. De Boer W, Besseling J, Willems J. Organisation of disability evaluation in 15 countries. Prat Organ Soins. 2007;3(38):205-217.
2. Vermeulen SJ, Tamminga SJ, Schellart AJ, Ybema JF, Anema JR. Return-to-work of sick-listed workers without an employment contract–what works? BMC Public Health. 2009;9(1):232.
3. UWV. Cijfers en trends UWV, januari 2020. 2020.
4. Lammerts L, Vermeulen SJ, Schaafsma FG, van Mechelen W, Anema JR. Return to work of workers without a permanent employment contract, sick-listed due to a common mental disorder: design of a randomised controlled trial. BMC Public Health. 2014;14(1):594.
5. Anner J, Schwegler U, Kunz R, Trezzini B, de Boer W. Evaluation of work disability and the international classification of functioning, disability and health: what to expect and what not. BMC Public Health. 2012;12(1):470.
6. Kok R, Hoving JL, Verbeek J, Schaafsma FG, van Dijk FJ. Integrating evidence in disability evaluation by social insurance physicians. Scand J Work Environ Health. 2011;37(6):494-501.
7. Kok R, Verbeek JA, Faber B, van Dijk FJ, Hoving JL. A search strategy to identify studies on the prognosis of work disability: a diagnostic test framework. BMJ Open. 2015;5(5):e006315.
8. Verbrugge LM, Jette AM. The disablement process. Soc Sci Med. 1994;38(1):1-14.
9. Mathiowetz N, Wunderlich GS. Conceptual issues in the measurement of work disability, in Survey measurement of work disability: Summary of a workshop, Mathiowetz NWunderlich G, Editors. 2000, National Academies Press: Washington.
10. Nagi SZ. Disability concepts revisited; implications for prevention, in Disability in America: Toward a national agenda for prevention, Pope ATarlow A, Editors. 1991, National Academy Press: Washington.
11. OECD. Sickness, Disability and Work: Breaking the Barriers. Paris: OECD Publishing; 2010.
12. Backhans MC, Hemmingsson T. Unemployment and mental health—who is (not) affected? Eur J Public Health. 2011;22(3):429-433.
13. Norström F, Virtanen P, Hammarström A, Gustafsson PE, Janlert U. How does unemployment affect self-assessed health? A systematic review focusing on subgroup effects. BMC Public Health. 2014;14(1):1310.
14. Milner A, LaMontagne A, Aitken Z, Bentley R, Kavanagh AM. Employment status and mental health among persons with and without a disability: evidence from an Australian cohort study. J Epidemiol Community Health. 2014;68(11):1064-1071.
15. Henderson M, Glozier N, Elliott KH. Long term sickness absence. 2005, British Medical Journal Publishing Group.
16. UWV. UWV Kennisverslag Volumeontwikkelingen 2019-I: Duiding van ontwikkelingen in de sociale zekerheid met speciale aandacht voor de gevolgen van het langer doorwerken. 2019.
17. Kerkhofs M, Lindeboom M, Theeuwes J. Retirement, financial incentives and health. Labour Econ. 1999;6(2):203-227.
18. Berecki-Gisolf J, Clay FJ, Collie A, McClure RJ. The impact of aging on work disability and return to work: insights from workers' compensation claim records. J Occup Environ Med. 2012;54(3):318-327.
19. Jansen H, Noordik E. Net geregistreerd als verzekeringsarts en dan weg bij UWV. Waarom? TBV. 2020;28(1-2):8-14.
20. Schepman S, Van der Velden L, Hingstman L. Capaciteitsraming voor bedrijfsartsen 2012-2022/2028. 2011, NIVEL: Utrecht.

21. Ilmarinen J. Work ability—a comprehensive concept for occupational health research and prevention. Scand J Work Environ Health. 2009;35(1):1-5.
22. Tuomi K, Ilmarinen J, Jahkola A, Katajarinne L, Tulkki A. Work Ability Index. Helsinki: Finnish Institute of Occupational Health; 1994.
23. De Zwart B, Frings-Dresen M, Van Duivenbooden J. Test–retest reliability of the Work Ability Index questionnaire. Occup Med. 2002;52(4):177-181.
24. Ilmarinen J. The work ability index (WAI). Occup Med. 2007;57(2):160-160.
25. Alavinia SM, De Boer A, Van Duivenbooden J, Frings-Dresen M, Burdorf A. Determinants of work ability and its predictive value for disability. Occup Med. 2008;59(1):32-37.
26. De Boer A, Verbeek J, Spelten E, Uitterhoeve A, Ansink A, De Reijke T, et al. Work ability and return-to-work in cancer patients. Br J Cancer. 2008;98(8):1342.
27. Sandqvist JL, Henriksson CM. Work functioning: A conceptual framework. Work. 2004;23(2):147-157.
28. Tengland PA. The concept of work ability. J Occup Rehabil. 2011;21(2):275-285.
29. Nieuwenhuijsen K, Verbeek J, De Boer A, Blonk RW, van Dijk FJ. Predicting the duration of sickness absence for patients with common mental disorders in occupational health care. Scand J Work Environ Health. 2006;32(1):67-74.
30. Holtermann A, Hansen JV, Burr H, Søgaard K. Prognostic factors for long-term sickness absence among employees with neck-shoulder and low-back pain. Scand J Work Environ Health. 2010;36(1):34-41.
31. Roelen CA, Koopmans PC, Groothoff JW, van der Klink JJ, Bültmann U. Sickness absence and full return to work after cancer: 2-year follow-up of register data for different cancer sites. Psycho-oncology. 2011;20(9):1001-1006.
32. Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, et al. Clinical decision support systems for the practice of evidence-based medicine. J Am Med Inform Assoc. 2001;8(6):527-534.
33. Bates DW, Cohen M, Leape LL, Overhage JM, Shabot MM, Sheridan T. Reducing the frequency of errors in medicine using information technology. J Am Med Inform Assoc. 2001;8(4):299-308.
34. Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux P, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA. 2005;293(10):1223-1238.
35. Roshanov PS, Fernandes N, Wilczynski JM, Hemens BJ, You JJ, Handler SM, et al. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. BMJ. 2013;346:f657.
36. Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J Am Med Inform Assoc. 2003;10(6):523-530.

CHAPTER 2

Characteristics of individuals receiving disability benefits in the Netherlands and predictors of leaving the disability benefit scheme: a retrospective cohort study with five-year follow-up

I Louwerse
MA Huysmans
HJ van Rijssen
AJ van der Beek
JR Anema

Published in:
BMC Public Health
2018;18(1):157

Abstract

Background Today, work disability is one of the greatest social and labour market challenges for policy makers in most OECD countries, where on average, about 6% of the working-age population relies on disability benefits. Understanding of factors associated with long-term work disability may be helpful to identify groups of individuals at risk for disability benefit entitlement or continuing eligibility, and to develop effective interventions for these groups. The purpose of this study is to provide insight into the main diagnoses of workers who qualify for disability benefits and how these diagnoses differ in age, gender and education. Using a five-year follow-up, we examined the duration of disability benefits and how durations differ among individuals with various characteristics.

Methods We performed a cohort study of 31,733 individuals receiving disability benefits from the Dutch Social Security Institute (SSI) with a five-year follow-up. Data were collected from SSI databases. Information about disorders was assessed by an insurance physician upon benefit application. These data were used to test for significant relationships among sociodemographics, main diagnoses and comorbidity, and disability benefit entitlement and continuing eligibility.

Results Mental disorders were the most frequent diagnosis for individuals claiming work disability. Diagnoses differed among age groups and education categories. Mental disorders were the main diagnosis for work disability for younger and more highly educated individuals, and physical disorders (generally musculoskeletal, cardiovascular and cancer) were the main diagnosis for older and less educated individuals. In 82% of the claims, the duration of disability benefit was five years or more after approval. Outflow was lowest for individuals with (multiple) mental disorders and those with comorbidity of mental and physical disorders, and highest for individuals with (multiple) physical disorders.

Conclusions The main diagnosis for persons entitled to disability benefits was mental health problems, especially for young women. In a five-year follow-up, claim duration for disability benefits was long lasting for most claimants.

Background

Today, work disability is one of the greatest social and labour market challenges for policy makers in most OECD countries [1]. On average, about 6% of the working-age population relies on disability benefits. Public spending on these benefits has become a serious burden. In the Netherlands, spending on disability benefits has risen to 4% of the gross domestic product. Moreover, once a disability benefit has been approved, the probability of returning to work is low [1]. Long-term unemployment or occupational inactivity is bad for an individual's health, especially those with mental health conditions, and returning to work is generally associated with health improvement [2-4]. Thus, prevention of work disability and support for returning to work are in the interest of individuals and society as a whole.

Dutch social security legislation allows workers to apply for a disability benefit after two years of sick leave (see text box) [5]. Disability benefits can be approved for a disease or handicap due to either social (i.e. non-occupational) or occupational causes.

In the Netherlands, disability benefits are assessed by the Dutch Social Security Institute (SSI). After two years of sick leave, individuals can apply for a disability benefit under the Dutch Work and Income Act (WIA). A medical disability assessment (i.e. diagnoses and functional abilities) is conducted by an insurance physician (IP) who is employed by the SSI. Depending on the functional abilities listed in the IP's report, there may also be an assessment by a labour expert who calculates the loss of former wages. A disability benefit is granted if loss of income exceeds 35% of former wages. Disability benefits can be approved for a disease or handicap due to either social (i.e. non-occupational) or occupational causes. Certain circumstances may change a person's continuing eligibility for disability benefits. A disability benefit ends if the SSI IP determines that the medical condition has improved substantially and the labour expert calculates that loss of income is less than 35% of former wages. Other main causes are retirement and death.

Return to work patterns and the effectiveness of interventions aimed at a return to work differ across individuals and their specific characteristics and health conditions [6, 7]. In Sweden, the majority of psychiatric outpatients with depression were female, less than 44 years, and had completed more than 9 years of compulsory education [8]. Moreover, the duration of mental health-related disability was influenced by sociodemographic factors such as age and education, and clinical factors such as comorbidity [9, 10]. A Dutch study among cancer survivors following 2 years of sick leave concluded that among other factors, higher education, physical limitations and low self-reported work ability were associated with an increased risk for work disability [11]. For other physical disorders, such as lower back pain and major limb trauma, older age, low education level and smoking were significant predictors for long-term work disability [12, 13].

Whereas most studies about individuals at risk for long-term disability benefits and factors affecting return to work focus on a specific diagnosis, an overview of all diagnoses is missing. Therefore, in the present study we included all individuals who had been granted a long-term disability benefit in the Netherlands. An overview of individual characteristics provides insight into the range of diagnoses and how they differ among age groups, gender and education. This approach can help to target specific at-risk groups and identify effective interventions to prevent long-term work disability. Therefore, the aim of this study was twofold. The first aim was to identify the most important diagnoses for which individuals claim an inability to work, and to examine how diagnoses differ among age groups, gender and educational levels. Second, using a five-year follow-up, we aimed to determine the duration of the disability benefit and how durations differed among individuals with various characteristics.

Methods

Study population

The study cohort included 31,733 subjects who had been granted a WIA disability benefit by the SSI between July 2010 and June 2011 after a medical disability assessment by an IP. Subjects in the study sample were assessed as having a full and permanent work disability, non-permanent but full work disability, or permanent and partial work disability. Individuals in the latter group had some work capacity and were possibly enrolled in a (part-time) job. Adults disabled since childhood were not included in the study sample since in the Netherlands they are not entitled to a WIA disability benefit (instead they can apply for a Disablement Assistance Act for Handicapped Young Persons disability benefit when they turn 18).

Sociodemographics

Sociodemographic data including gender, age and education are registered in the SSI database upon application for benefits. For further analysis, age was categorized into four groups: <35, 35-44, 45-54 and 55+ years. Three education levels were defined based on the highest level of education completed; low (primary school, lower vocational education, lower secondary school), secondary (intermediate vocational education, upper secondary school), and high (upper vocational education, university). The educational level is usually registered during the labour expert's assessment. Since this assessment is not necessary when the IP assesses full and permanent work disability, the education level was missing for 4,036 individuals in our study sample. We excluded these individuals from the analyses concerning the education level, as we could not deduce any information about their educational level and were therefore not able to use the results.

Disorders

When applying for a disability benefit, the assessment of diagnoses and functional abilities is done by an IP who is employed by the SSI. The IP lists disorders according to the Dutch Classification of Occupational Health and Social Insurance (CAS). The CAS is based on the International Statistical Classification of Disease and Related Health Problems (ICD-10) diseases, a medical classification list from the World Health Organization [14]. The IP can list up to three disorders during the medical disability assessment. These diagnoses are divided into 14 categories, according to the ICD-10 classification, which we used in this study.

Comorbidity

For this study, we created a comorbidity classification scheme based on the CAS, as established classification schemes did not fit our study data. CAS includes only information about the existence of disorders, and not about their severity. The IP lists in CAS the first (main) diagnosis for which an individual claims inability to work, and possibly a second and third diagnosis. The IP will only mention a second or third diagnoses if he or she believes that these result in important, additional functional disabilities. Therefore, in the present study, we considered all second and third diagnoses as comorbidity, independent of the disease categories of the ICD-10 classification that the disorders belong to.

We defined comorbidity as two or three disorders being listed for an individual. To gain insight into the disorders present in cases of comorbidity, we divided the diagnostic categories into mental (mental disorders) and physical disorders (all remaining disorders). Possible conditions of comorbidity were multiple mental disorders, multiple physical disorders or a comorbidity of mental and physical disorders.

Continuing eligibility for disability benefit

We used a follow-up period of five years. For each individual in the study sample, we used SSI registration data to determine whether the benefit ended within one, three or five years after the date of approval (and if so, for what reason). During this five-year follow-up period, there were no major changes in legislation or working processes that could have influenced our results.

Statistical Analysis

Statistical analyses were performed in RStudio for Windows, version 0.99.902. The chi-square test for categorical variables was used to compare sociodemographic characteristics, disorders, comorbidity and outflow from disability benefits among various groups of individuals.

Multinomial regression models were used to test for relationships between disorders, comorbidity and outflow from disability benefit respectively (dependent variable) and socio-demographic characteristics, disorders and comorbidity (independent variables) while taking confounding effects into account. The level of significance was set at $p<0.05$.

Results

Characteristics of the study population

The sociodemographic characteristics and disorders of the study population are summarized in table 1. To facilitate interpretation, all numbers were rounded to the nearest ten. The mean age was 46.8 years (SD, 10.6) and the number of men and women was approximately equal. Table 2 shows the age categories and educational levels divided by gender. Women who qualified for disability benefits were on average younger [χ^2(df=3; n=31,733)=519.33, p=0.000], and more highly educated [χ^2(df=2; n=27,697)=262.43, p=0.000] than men.

Table 1. Summary of sociodemographic characteristics and disorders (n=31,730)

	Study sample, n (%)
Sociodemographics	
Gender	
Male	15,650 (49)
Female	16,090 (51)
Age category	
< 35	5,210 (16)
35-44	7,060 (22)
45-54	10,160 (32)
55+	9,300 (29)
Educational level	
Low	16,820 (53)
Secondary	7,390 (23)
High	3,500 (11)
Unknown	4,040 (13)
Disorders	
Main causes of work disability	
Cancer	2,510 (8)
Cardiovascular	2,790 (9)
Mental	10,870 (34)
Musculoskeletal	8,410 (27)
Nervous system	2,090 (9)
Other	4,260 (13)

Table 2. Age and educational level by gender

	Gender, n (%)	
	Male	Female
Age category		
< 35	2,070 (13)	3,140 (20)
35-44	3,230 (21)	3,830 (24)
45-54	4,940 (32)	5,220 (33)
55+	5,400 (35)	3,890 (24)
Educational level		
Low	8,860 (57)	7,960 (50)
Secondary	3,290 (21)	4,100 (26)
High	1,400 (9)	2,100 (13)
Unknown	2,100 (13)	1,940 (12)

Table 3. Main diagnosis by age, gender and educational level

	Main diagnosis, n (%)					
	Mental	Musculo-skeletal	Nervous system	Cardio-vascular	Cancer	Other
Gender						
Male	5,160 (33)	4,170 (27)	1,460 (9)	1,860 (12)	890 (6)	2,110 (13)
Female	5,710 (35)	4,240 (26)	1,450 (9)	930 (6)	1,620 (10)	2,150 (13)
Age category						
< 35	2,970 (57)	1,000 (19)	460 (9)	80 (1)	120 (2)	590 (11)
35-44	3,150 (45)	1,690 (24)	670 (10)	320 (5)	360 (5)	860 (12)
45-54	3,010 (30)	2,910 (29)	950 (9)	1,010 (10)	960 (9)	1,340 (13)
55+	1,740 (19)	2,810 (30)	820 (9)	1,380 (15)	1,070 (12)	1,470 (16)
Educational level						
Low	5,350 (32)	5,860 (35)	1,210 (7)	1,470 (9)	830 (5)	2,100 (12)
Secondary	2,680 (36)	1,820 (25)	750 (10)	610 (8)	470 (6)	1,070 (14)
High	1,510 (43)	470 (13)	450 (13)	280 (8)	280 (8)	500 (14)

Main diagnoses for work disability

Table 1 shows the main disorders as listed by the IP for medical disability assessment. Mental disorders were most often mentioned, followed by musculoskeletal disorders, nervous system disorders, cancer and cardiovascular system disorders. The category "other" consisted of various classes of physical disorders that were listed less frequently (among others respiratory system, digestive system and genitourinary system disorders).

Table 3 shows that the main diagnosis for work disability differed significantly among age categories [χ^2(df=15; n=31,733)=3306.1, p=0.000], with mental disorders as the main diagnosis for individuals younger than 55 years, and musculoskeletal disorders the main diagnosis for individuals 55 years and older. The differences in leading diagnoses between men and women were statistically significant but smaller [χ^2(df=5; n=31,733)=541.15, p=0.000]. Mental and musculoskeletal disorders were registered with approximately the same frequency for women and men. However, cancer was more often registered for women (mostly breast cancer) and cardiovascular disorders for men (mostly stroke, heart attack).

The leading diagnosis also differed for educational level [χ^2(df=10; n=27,697)=857.12, p=0.000]. Individuals who were more highly educated suffered more often from mental disorders, nervous system disorders and cancer, whereas individuals who were less educated suffered more often from musculoskeletal and cardiovascular disorders.

Comorbidity

More than half of the individuals in the study population (55.8%) suffered from comorbidity. Table 4 shows that work disability due to comorbidity was mentioned as frequently for men as for women [χ^2(df=1; n=31,733)=0.356, p=0.551], and more often for older [χ^2(df=3; n=31,733)=92.866, p=0.000] and less educated individuals [χ^2(df=2; n=27,697)=168.65, p=0.000]. Considering the main diagnoses for work disability [χ^2(df=5; n=31,733)=765.29, p=0.000], individuals with cancer suffered least often from comorbidity and individuals with musculoskeletal disorders most often.

Table 4. Comorbidity by sociodemographics and main diagnosis

	Comorbidity, n (%)	
	Yes	No
Gender		
Male	8,700 (56)	6,940 (44)
Female	9,010 (56)	7,080 (44)
Age category		
< 35	2,690 (52)	2,520 (48)
35-44	3,770 (53)	3,290 (47)
45-54	5,780 (57)	4,390 (43)
55+	5,470 (59)	3,830 (41)
Educational level		
Low	10,240 (61)	6,580 (39)
Secondary	4,060 (55)	3,330 (45)
High	1,760 (50)	1,730 (50)
Main diagnosis causing work disability		
Cancer	930 (37)	1,580 (63)
Cardiovascular	1,640 (59)	1,150 (41)
Mental	6,310 (58)	4,560 (42)
Musculoskeletal	5,180 (62)	3,230 (38)
Nervous system	1,190 (41)	1,710 (59)
Other	2,460 (58)	1,800 (41)

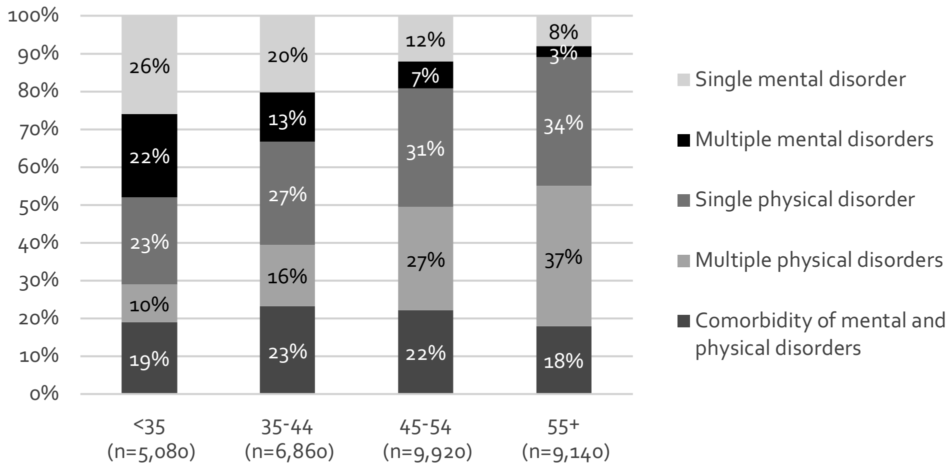


Figure 1. Comorbidity for individuals of various age groups

Figure 1 shows the (combination of types of) diagnoses registered by the IP during a medical disability assessment for each age category. For younger individuals, comorbidity was most often a combination of multiple mental disorders, and for older individuals it was most often a combination of multiple physical diagnoses (musculoskeletal, nervous or cardiovascular disorders).

Continuing eligibility for disability benefits

Of the individuals who had been granted a disability benefit in the study time frame, 964 individuals (3%) left in the first year, 2,607 (8%) in the second and third year, and 2,258 (7%) in the fourth and fifth year. All other individuals (25,907; 82%) continued to receive disability benefits after five years. Outflow was caused by retirement (37%), death (30%), improvement of the medical condition (such that individuals could earn at least 65% of former wages [28%]), or other reasons such as imprisonment or pregnancy (5%).

Table 5 shows that the main reason for outflow in the first year was death, in the second and third year income loss lower than 35% of former wages earned, and in the fourth and fifth year retirement. Table 6 shows that the differences for gender [χ^2(df=3; n=31,733)=37.549, p=0.000] and age categories [χ^2(df=9; n=31,733)=2259.8, p=0.000] were statistically significant, but small. There was no difference for education categories [χ^2(df=6; n=27,697)=5.9007, p=0.434].

The outflow of disability benefits differed by class of the leading diagnoses for work disability [χ^2(df=15; n=31,733)=2148.4, p=0.000]. In the first year, the outflow consisted mainly of individuals diagnosed with cancer who died within one year after their disability benefit was approved. After four and five years, more older individuals with musculoskeletal and cardio-vascular disorders left because of retirement.

Continuing eligibility for disability benefits was highest for individuals with single or multiple mental disorders and individuals facing a comorbidity of mental and physical disorders, and lowest for individuals with single or multiple physical disorders [χ^2(df=15; n=1,733)=653.9, p=0.000].

Table 5. Timing and reason for outflow of disability benefit

	Reason for outflow of disability benefit, n (%)				
	Retirement	Death	Income loss <35%	Other	Total
Outflow of disability benefit					
1st year	200 (21)	490 (52)	170 (18)	100 (10)	960 (100)
2nd or 3rd year	790 (30)	750 (29)	930 (36)	140 (5)	2,610 (100)
4th or 5th year	1,140 (50)	510 (22)	540 (24)	80 (3)	2,260 (100)

Table 6. Outflow of disability benefits by sociodemographics, main diagnosis and comorbidity

	Outflow of disability benefit, n (%)			
	1st year	2nd or 3rd year	4th or 5th year	Continuing eligibility
Gender				
Male	510 (3)	1,330 (8)	1,230 (8)	12,590 (80)
Female	450 (3)	1,280 (8)	1,030 (6)	13,330 (83)
Age category				
< 35	110 (2)	370 (7)	200 (4)	4,530 (87)
35-44	130 (2)	510 (7)	220 (3)	6,200 (88)
45-54	230 (2)	560 (5)	380 (4)	9,000 (88)
55+	490 (5)	1,170 (13)	1,450 (16)	6,180 (66)
Education level				
Low	360 (2)	1,300 (8)	1,200 (7)	13,950 (83)
Secondary	160 (2)	570 (8)	450 (7)	6,150 (83)
High	70 (2)	290 (8)	270 (8)	2,860 (82)
Main diagnosis causing work disability				
Mental	160 (1)	670 (6)	540 (5)	9,500 (87)
Musculoskeletal	190 (2)	650 (8)	570 (7)	7,000 (83)
Nervous system	50 (2)	150 (5)	220 (7)	2,490 (86)
Cardiovascular	60 (2)	250 (9)	300 (11)	2,180 (78)
Cancer	370 (15)	490 (19)	250 (10)	1,410 (56)
Other	140 (3)	400 (9)	390 (9)	3,330 (78)
Comorbidity				
Single mental disorder	80 (2)	310 (7)	220 (5)	4,020 (87)
Multiple mental disorders	30 (1)	160 (5)	130 (4)	2,680 (89)
Single physical disorder	490 (5)	940 (10)	770 (8)	7,070 (76)
Multiple physical disorders	250 (3)	740 (10)	710 (9)	6,010 (78)
Mental and physical disorders	100 (2)	410 (6)	390 (6)	5,490 (86)

Discussion

Main findings

Disability diagnoses differed significantly among age groups and education categories; mental disorders were the main diagnosis for work disability for younger and more highly educated individuals, and physical disorders (mainly musculoskeletal, cardiovascular and cancer) for older and less educated individuals. The differences between men and women were small. Multiple diagnoses were registered for more than half of the population. Older and less educated individuals suffered relatively often from comorbidity. In the five-year follow-up, the continuation of disability benefits for five years or more after approval was high. Only 18% of the individuals in our study sample discontinued their disability benefits in the five-year follow-up period. Continuing eligibility for disability benefit was highest for individuals with (multiple) mental disorders and those with a comorbidity of mental and physical disorders, and lowest for individuals with (multiple) physical disorders.

Interpretation of findings and comparison with other studies

The current finding that women who qualify for disability benefits are on average younger and more educated than men confirms previously reported findings. A reason for this difference in age of entry to disability benefits is the relatively low number of older women among the insured population [15]. This is most likely because a few decades ago many women did not continue paid work after giving birth. More recently, the proportion of women aged 50 -64 years in the workforce has increased, and is still increasing such that the employment gap between men and women is becoming smaller [16,17].

We found that mental disorders were the main diagnosis for work disability. This is in line with the finding that mental disorders are the leading cause of sickness absence and work disability in OECD countries [18]. Research shows that mental health impairments have increased over the past years. This could be explained by the changing content of communication and social networks, and the changed and increased job demands in the workplace [19, 20]. All these factors make it increasingly difficult for workers with mental health problems to return to work. Our finding that younger individuals in particular suffer from mental disorders corresponds with the finding that younger generations are at increased risk for mental health problems [1, 21]. Two major explanations are changes in the workplace that have increased the prevalence of work-related stress, and the changing content of communication and social networks. Our finding is a problem because work absence due to mental illness is often long lasting. In the Netherlands, the median duration of absence due to mental illness has increased. The probability of resuming work decreases with the increasing duration of absence due to illness [22]. Conversely, the prevalence of musculoskeletal disorders as the main diagnosis for work disability is higher for older individuals. An association between age and musculoskeletal disorders is generally found in several studies [23, 24].

Concerning the relationship between the main diagnoses for work disability and education, we found that individuals who were more highly educated suffered more often from mental disorders, nervous system disorders and cancer, and individuals who were less educated suffered more often from musculoskeletal and cardiovascular disorders. This may be due to differences in the type of jobs and workplaces for these two groups.

A considerable part of our study population (65%) suffered from comorbidity. Research shows the importance of comorbidity as a predictor for long-term work disability [25]. Multiple physical symptoms have a generic negative influence on the effectiveness of treatment for symptoms of depression and anxiety in primary care [26].

The duration of disability benefits is longer for older workers, when the main diagnosis for work disability is a mental disorder and when comorbidity is present, and only related to gender and education to a limited extent. Similar findings can be found in the literature on prognostic factors for long-term disability due to mental disorders [27]. Of the individuals in the study population, 82% had continuing eligibility for their disability benefits five years after approval. An application for disability benefits can be requested after two years of sick leave. This means that individuals who qualify for disability benefits have already been sick for a long period of time and have severe disorders that may be more difficult to treat. In addition, in these two years, the system does not offer many incentives for individuals to return to work. Hence, (partial) recovery after two years of sick leave would be unexpected. This could explain the low outflow in the present cohort.

Strengths and limitations of the study

An important strength of this study is the large study sample. By covering the entire Dutch population applying for long-term disability benefits, with a one-year inflow period and a five-year follow-up period, our study population is highly inclusive. To our knowledge studies about individuals at risk for long-term disability benefits generally focus on one specific diagnosis, while we included all individuals who were granted a disability benefit in the Netherlands in the one-year inflow period. By doing this, we can give an overview of all diagnoses for which individuals claim work disability. In addition, in most studies in the field of work disability the follow-up period is only one or two years, while we were able to use a follow-up period of five years after approval of the benefit. We performed a similar study with individuals who were granted a disability benefit in 2015 and the individual characteristics, main diagnoses for work disability and comorbidity numbers were similar to the ones in this study, thus confirming our results here. The figures on the sociodemographic characteristics of individuals receiving disability benefits are also consistent with SSI data [28]. A limitation of only testing for bivariate relationships is that it is not possible to control for confounding effects. Therefore, we have also performed three multinomial regression analyses (with main diagnosis for work disability, comorbidity and continuing eligibility for disability benefits respectively as the dependent variables). The results of these regression analyses can be found in Appendix A. They confirm the statistical bivariate relations that we found with the chi-square tests.

A study limitation is that data was not collected for research purposes, but rather registered by SSI employees for administration purposes. Although careful registration is important for internal processes, employees might not have been fully aware of the importance of complete and comprehensive administration and some records contained missing data. For that reason, we had to exclude 4,036 individuals from our analyses concerning the education level as their values were missing. In this study, we considered only sociodemographic factors, main diagnosis, comorbidity and claim duration. However, there could be other factors (partly) explaining our findings.

Practical implications

This study provides insight into the sociodemographic factors and health complaints of individuals who qualify for disability benefits in the Netherlands and shows that continuing eligibility for disability benefits is high. This information can help identify specific at-risk groups when policies are aimed at decreasing the number of applications for disability benefits. The results of this study may be useful when policy makers investigate how to reduce long-term disability benefits. In this context, the focus should be on individuals who leave for reasons other than retirement and death. Increased understanding of the characteristics of this group and how to support them in returning to work is needed. Conducting re-assessments, wherein the SSI would assess whether or not someone's health had improved enough so that their earning capacity had increased, is a possible way to motivate individuals to return to work.

Conclusions

This study provides an overview of the sociodemographic characteristics and diagnoses of individuals who have been granted a disability benefit, and examines the duration of their benefit. Therefore, it contributes to insight into the range of diagnoses and how they differ in age, gender and education. An understanding of factors associated with long-term work disability may be helpful to identify groups of individuals who are at risk for continuing eligibility for disability benefits and to develop effective interventions for these groups.

Appendix

Table 7. Relation between main diagnosis for work disability and sociodemographic characteristics

Main diagnosis[a]	Independent variables	Coefficient	Standard error	*Z* statistic	*p*-value	RRR[b]
Musculoskeletal	Intercept	-2.484	0.074	-33.47	0.000	
	Gender					
	Male	1.00 (ref)				
	Female	0.095	0.031	3.10	0.002	1.10
	Age	0.055	0.001	37.17	0.000	1.06
	Educational level					
	Low	1.00 (ref)				
	Secondary	-0.379	0.039	-10.27	0.000	0.68
	High	-1.293	0.057	-22.59	0.000	0.27
	Unknown	-1.820	0.072	-25.41	0.000	0.16
Nervous system	Intercept	-3.135	0.104	-30.15	0.000	
	Gender					
	Male	1.00 (ref)				
	Female	-0.018	0.043	-0.43	0.666	0.98
	Age	0.037	0.002	28.23	0.000	1.04
	Educational level					
	Low	1.00 (ref)				
	Secondary	0.284	0.053	5.37	0.000	1.33
	High	0.260	0.063	4.11	0.000	1.30
	Unknown	0.447	0.062	7.21	0.000	1.56
Cardiovascular	Intercept	-6.185	0.141	-43.97	0.000	
	Gender					
	Male	1.00 (ref)				
	Female	-0.518	0.046	-11.17	0.000	0.60
	Age	0.106	0.003	40.37	0.000	1.11
	Educational level					
	Low	1.00 (ref)				
	Secondary	0.014	0.056	0.25	0.805	1.01
	High	-0.395	0.074	-5.33	0.000	0.67
	Unknown	-0.001	0.066	-0.02	0.985	1.00
Cancer	Intercept	-6.615	0.144	-45.87	0.000	
	Gender					
	Male	1.00 (ref)				
	Female	0.775	0.049	15.92	0.000	2.17
	Age	0.091	0.003	34.93	0.000	1.10
	Educational level					
	Low	1.00 (ref)				
	Secondary	0.257	0.064	4.00	0.000	1.29
	High	0.096	0.077	1.28	0.200	1.10
	Unknown	1.396	0.059	23.52	0.000	4.04
Other	Intercept	-3.303	0.093	-35.48	0.000	
	Gender					
	Male	1.00 (ref)				
	Female	0.061	0.037	1.63	0.103	1.06
	Age	0.051	0.002	28.23	0.000	1.05
	Educational level					
	Low	1.00 (ref)				
	Secondary	0.107	0.045	2.36	0.018	1.11
	High	-0.197	0.059	-3.36	0.000	0.82
	Unknown	0.035	0.057	0.61	0.543	1.04

[a] Reference category is mental disorders, [b] RRR = relative risk ratio

Table 8. Relation between comorbidity and sociodemographic characteristics and main diagnosis for work disability

Comorbidity[a]	Independent variables	Coefficient	Standard error	*Z* statistic	*p*-value	RRR[b]
Multiple mental, multiple physical or comorbidity of mental and physical disorders	Intercept	-0.231	0.057	-4.08	0.000	
	Gender					
	Male	1.00 (ref)				
	Female	0.103	0.024	4.37	0.000	0.80
	Age	0.016	0.001	13.99	0.000	1.11
	Educational level					
	Low	1.00 (ref)				
	Secondary	-0.193	0.029	-6.71	0.000	0.82
	High	-0.382	0.038	-9.96	0.000	0.68
	Unknown	-0.670	0.037	-17.88	0.000	0.51
	Main diagnosis					
	Mental	1.00 (ref)				
	Musculoskeletal	-0.044	0.031	-1.40	0.160	0.96
	Nervous system	-0.723	0.043	-16.73	0.000	0.49
	Cardiovascular	-0.121	0.045	-2.66	0.008	0.89
	Cancer	-0.887	0.048	-18.41	0.000	0.41
	Other	-0.095	0.038	-2.53	0.013	0.91

[a] Reference category is no comorbid disorders (single mental or single physical disorder)
[b] RRR = relative risk ratio

Table 9. Relation between continuing eligibility for disability benefit and sociodemographic characteristics, main diagnosis for work disability and comorbidity

Outflow of benefit[a]	Independent variables	Coefficient	Standard error	*Z* statistic	*p*-value	RRR[b]
Outflow in 1st year	Intercept	-5.934	0.230	-25.78	0.000	
	Gender					
	Male	1.00 (ref)				
	Female	-0.255	0.070	-3.64	0.000	0.78
	Age	0.042	0.004	10.36	0.000	1.04
	Educational level					
	Low	1.00 (ref)				
	Secondary	0.057	0.099	0.58	0.564	1.06
	High	-0.024	0.133	-0.18	0.860	0.98
	Unknown	1.033	0.087	11.84	0.000	2.81
	Main diagnosis					
	Mental	1.00 (ref)				
	Musculoskeletal	-0.347	0.214	-1.62	0.105	0.71
	Nervous system	-0.925	0.249	-3.71	0.000	0.40
	Cardiovascular	-0.761	0.241	-3.15	0.002	0.47
	Cancer	1.351	0.213	6.33	0.000	3.86
	Other	-0.079	0.221	-0.17	0.864	0.96
	Comorbidity					
	Single mental disorder	1.00 (ref)				

2

Table 9. Continued

Outflow of benefit[a]	Independent variables	Coefficient	Standard error	Z statistic	p-value	RRR[b]
	Multiple mental disorders	-0.302	0.211	-1.43	0.153	0.74
	Single physical disorder	0.876	0.242	3.62	0.000	2.10
	Multiple physical disorders	0.617	0.246	2.51	0.012	1.85
	Mental and physical	-0.066	0.184	-0.36	0.721	0.94
Outflow in 2nd or 3rd year	Intercept	-4.060	0.126	-32.15	0.000	
	Gender					
	Male	1.00 (ref)				
	Female	-0.079	0.043	-1.83	0.067	0.92
	Age	0.034	0.002	14.57	0.000	1.03
	Educational level					
	Low	1.00 (ref)				
	Secondary	0.068	0.054	1.27	0.203	1.07
	High	0.088	0.070	1.25	0.210	1.09
	Unknown	0.204	0.063	3.21	0.001	1.22
	Main diagnosis					
	Mental	1.00 (ref)				
	Musculoskeletal	-0.114	0.108	-1.62	0.293	0.89
	Nervous system	-0.572	0.133	-4.32	0.000	0.56
	Cardiovascular	-0.098	0.122	-0.81	0.421	0.91
	Cancer	1.029	0.116	8.85	0.000	2.80
	Other	0.184	0.115	1.60	0.110	1.20
	Comorbidity					
	Single mental disorder	1.00 (ref)				
	Multiple mental disorders	-0.133	0.100	-1.32	0.186	0.88
	Single physical disorder	0.238	0.125	1.90	0.058	1.27
	Multiple physical disorders	0.171	0.127	1.35	0.178	1.19
	Mental and physical	-0.180	0.095	-1.89	0.880	0.84
Outflow in 4th or 5th year	Intercept	-6.879	0.170	-40.54	0.000	
	Gender					
	Male	1.00 (ref)				
	Female	-0.048	0.046	-1.04	0.300	0.95
	Age	0.087	0.003	28.97	0.000	1.09
	Educational level					
	Low	1.00 (ref)				
	Secondary	0.086	0.057	1.50	0.133	1.09
	High	0.103	0.073	1.42	0.157	1.11
	Unknown	-0.142	0.074	-1.92	0.055	0.87
	Main diagnosis					
	Mental	1.00 (ref)				
	Musculoskeletal	-0.254	0.112	-2.27	0.023	0.78
	Nervous system	-0.102	0.127	-0.80	0.422	0.90
	Cardiovascular	-0.035	0.122	-0.29	0.772	0.97
	Cancer	0.371	0.127	2.92	0.004	1.45
	Other	0.193	0.118	1.64	0.101	1.21

Table 9. Continued

Outflow of benefit[a]	Independent variables	Coefficient	Standard error	*Z* statistic	*p*-value	RRR[b]
	Comorbidity					
	Single mental disorder	1.00 (ref)				
	Multiple mental disorders	0.073	0.117	0.62	0.535	1.08
	Single physical disorder	0.295	0.134	2.20	0.028	1.34
	Multiple physical disorders	0.195	0.135	1.44	0.150	1.21
	Mental and physical	-0.016	0.104	-0.15	0.880	0.98

[a] Reference category is continuing eligibility for disability benefit
[b] RRR = relative risk ratio

References

1. OECD. Sickness, Disability and Work: Breaking the Barriers: A Synthesis of Findings across OECD Countries. OECD Publishing. 2010. doi:10.1787/9789264088856-en
2. Norström F, Virtanen P, Hammarström A, Gustafsson PE, Janlert U. How does unemployment affect self-assessed health? A systematic review focusing on subgroup effects. BMC Public Health. 2014;14:1310
3. Backhans MC, Hemmingsson T. Unemployment and mental health—who is (not) affected?. Eur J Public Health. 2011;22:429-433.
4. Milner A, LaMontagne AD, Aitken Z, Bentley R, Kavanagh AM. Employment status and mental health among persons with and without a disability: evidence from an Australian cohort study. J Epidemiol Community Health. 2014;68:1064-1071.
5. OECD. Sickness and Disability Schemes in the Netherlands. http://www.oecd.org/social/soc /41429917.pdf. Accessed 21 Feb 2017.
6. Steenstra IA, Knol DL, Bongers PM, Anema JR, van Mechelen W, de Vet HC. What Works Best for Whom?: An Exploratory, Subgroup Analysis in a Randomized, Controlled Trial on the Effectiveness of a Workplace Intervention in Low Back Pain Patients on Return to Work. Spine. 2009;34:1243-1249.
7. Brouwer S, Reneman MF, Bültmann U, van der Klink JJL, Groothoff JW. A Prospective Study of Return to Work Across Health Conditions: Perceived Work Attitude, Self-efficacy and Perceived Social Support. J Occup Rehabil. 2010;20:104-112.
8. Mittendorfer-Rutz E, Härkänen T, Tiihonen J, Haukka J. Association of socio-demographic factors, sick-leave and health care patterns with the risk of being granted a disability pension among psychiatric outpatients with depression. PloS one. 2014;9, e99869
9. Lagerveld SE, Bültmann U, Franche RL, Van Dijk FJH, Vlasveld MC, Van der Feltz-Cornelis CM, et al. Factors associated with work participation and work functioning in depressed workers: a systematic review. J Occup Rehabil. 2010;20:275-292.
10. Ervasti J, Vahtera J, Pentti J, Oksanen T, Ahola K, Kivimäki M., Virtanen M. Depression-related work disability: socioeconomic inequalities in onset, duration and recurrence. PloS one. 2013;8: e79855.
11. Van Muijen P, Duijts SF, Bonefaas-Groenewoud K, van der Beek AJ, Anema JR. Factors associated with work disability in employed cancer survivors at 24-month sick leave. BMC cancer. 2014;14:236.
12. MacKenzie EJ, Bosse MJ, Kellam JF, Pollak AN, Webb LX, Swiontkowski MF, et al. Early predictors of long-term work disability after major limb trauma. J Trauma, 2006:61;688-694.
13. Natvig B, Eriksen W, Bruusgaard D. Low back pain as a predictor of long-term work disability. Scand J Public Health. 2002;30:288-292
14. World Health Organization: International Classification of Diseases. http://www.who.int /classifications/icd/en/ (2016). Accessed 6 Feb 2017.
15. Payne S, Doyal L. Older women, work and health. Occup Med (Lond). 2010;60:172-7.
16. Najati Jr G, Lee S, McCann D. Conditions of work and employment for older workers in industrialized countries: understanding the issues. No. 15. International Labour Organization. 2006.
17. Dixon S. Implications of population ageing for the labour market. Labour Mark Trends. 2003;3:67-76.
18. OECD. Sick on the Job?: Myths and Realities about Mental Health and Work. OECD Publishing. 2012. doi:10.1787/9789264124523-en.
19. Olfson M, Druss BG, Marcus SC. Trends in Mental Health Care among Children and Adolescents. N Engl J Med. 2015;372:2029-2038.
20. Twenge JM. Time Period and Birth Cohort Differences in Depressive Symptoms in the U.S., 1982-2013. Soc Indic Res. 2015;121:437-454.

21. Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, et al. Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. Acta psychiatrica scandinavica 2004;109:21-27.
22. Van Hoffen MFA, Joling CI, Heymans MW, Twisk JWR, Roelen CAM. Mental health symptoms identify workers at risk of long-term sickness absence due to mental disorders: prospective cohort study with 2-year follow-up. BMC Public Health. 2015;15:1235.
23. Kinge JM, Knudsen AK, Skirbekk V, Vollset SE. Musculoskeletal disorders in Norway: prevalence of chronicity and use of primary and specialist health care services. BMC Musculoskeletal Disorders. 2015;16:75.
24. Woolf AD, Pfleger B. Burden of major musculoskeletal conditions. Bulletin of the World Health Organization 2013;81: 646-656.
25. Ervasti J, Kivimäki M, Pentti J, Salo P, Oksanen T, Vahtera J, Virtanen M. Health- and work-related predictors of work disability among employees with a cardiometabolic disease—A cohort study. J Psychosomatic Research. 2016;82:41-47.
26. Sorvaniemia M, Heleniusb H, Salokangas RKR. Factors associated with being granted a pension among psychiatric outpatients with major depression. J Affect Disord. 2003;75:43-48.
27. Cornelius LR, Van der Klink JJL, Groothoff JW, Brouwer S. Prognostic factors of long term disability due to mental disorders: a systematic review. J Occup Rehabil. 2011;21:259-274.
28. Dutch Social Security Agency. Annual statistical report on the Security Disability Insurance Program. http://www.uwv.nl/overuwv/Images/UWV%20Kennisverslag%202015-2.pdf Accessed 6 Feb 2017.

CHAPTER 3

Predicting long-term sickness absence and identifying subgroups among individuals without an employment contract

I Louwerse
HJ van Rijssen
MA Huysmans
AJ van der Beek
JR Anema

Published in:
Journal of Occupational Rehabilitation
2020;30(3):371-380

Abstract

Purpose Today, decreasing numbers of workers in Europe are employed in standard employment relationships. Temporary contracts and job insecurity have become more common. This study among workers without an employment contract aimed to (i) predict risk of long-term sickness absence and (ii) identify distinct subgroups of sick-listed workers.

Methods 437 individuals without an employment contract who were granted a sickness absence benefit for at least two weeks were followed for one year. We used registration data and self-reported questionnaires on sociodemographics, work-related, health-related and psychosocial factors. Both were retrieved from the databases of the Dutch Social Security Institute and measured at the time of entry into the benefit. We used logistic regression analysis to identify individuals at risk of long-term sickness absence. Latent class analysis was used to identify homogenous subgroups of individuals.

Results Almost one-third of the study population (n=133; 30%) was still at sickness absence at one-year follow-up. The final prediction model showed fair discrimination between individuals with and without long-term sickness absence (optimism adjusted AUC to correct for overfitting =0.761). Four subgroups of individuals were identified based on predicted risk of long-term sickness absence, self-reported expectations about recovery and return to work, reason of sickness absence and coping skills.

Conclusion The logistic regression model could be used to identify individuals at risk of long-term sickness absence. Identification of risk groups can aid professionals to offer tailored return to work interventions.

Introduction

There is a positive association between work and one's well-being, mental and physical health [1, 2]. In contrast, unemployment is strongly associated with poor health. The longer individuals are absent from work, the less likely they are to return [3-5]. Although long-term sickness absence makes up only a relatively small proportion of absences, it accounts for more than one-third of days off and 75% of sickness absence costs [6]. Early identification of individuals at risk of long-term sickness absence and an overview of factors associated with sickness absence duration can help occupational health professionals to target specific at-risk groups and identify effective early interventions to prevent long-term sickness absence [7]. Because occupational health services resources are limited, a differentiated approach is needed in occupational rehabilitation offering different levels of return to work support depending on individual characteristics and needs. Identification of groups of individuals, which are similar on certain characteristics, could be used as a triage tool to identify groups of claimants with the highest risk of long-term sickness absence and offer them suitable return to work interventions, based on the group characteristics.

Today, decreasing numbers of workers in Europe are employed in standard employment relationships. Temporary contracts and job insecurity have become more common [8]. Workers without a permanent employment contract, i.e. unemployed and temporary agency workers, represent a vulnerable group within the working population as they have poorer health status, and increased risk of long-term sickness absence and work disability [9, 10]. They have a greater distance to the labour market as they are characterised by lower credentials, lower income, more females, more (partly) disabled, and more immigrants [11]. The biopsychosocial model of illness and disability proposes that return to work of sick-listed workers depends on a combination of biological, psychological and social factors [12]. As not having a permanent employment contract has a negative impact on the development and maintenance of psychosocial health, the interaction between the factors of the biopsychosocial model is different between workers with and without a permanent employment contract [13]. Furthermore, the fact that workers without a permanent employment contract usually do not have a workplace to return to, might complicate their return to work process and prolong their sickness absence duration. In the Netherlands, this is reflected in a higher number of workers still being sick-listed at one-year follow-up than workers with a permanent employment contract, and a higher inflow into work disability benefits after two years of sickness absence [14].

However, most studies on prognostic factors for long-term sickness absence focus on sick-listed employees, i.e. sick-listed workers with a permanent employment contract, rather than sick-listed workers without a permanent employment contract. Moreover, these studies focus on individuals with specific characteristics, for instance on individuals with specific diagnoses such as mental health problems [15-19], musculoskeletal disorders [20-23], or cancer [24, 25], or individuals belonging to a certain occupational group, such as healthcare workers [26]. These studies showed that sickness absence duration is mostly determined by factors that are not disorder-related. Although for occupational health professionals a prediction model that could be used for all diagnoses and occupational groups would be useful, such a model is currently missing.

In the present study, we included unemployed workers, temporary agency workers and workers with an expired fixed-term contract who received a sickness absence benefit for at least two weeks, covering all diagnoses and occupational groups. The aims of this study were to (i) predict sickness absence at one-year follow-up and (ii) explore whether distinct subgroups of sick-listed workers could be identified, partly based on their predicted risk of long-term sickness absence.

Methods

Study population

Dutch social security legislation allows sick-listed workers without a permanent employment contract to apply for a sickness absence benefit at the Dutch Social Security Institute (SSI; see text box) [27]. The study cohort included individuals who had been granted a sickness absence benefit by two regional offices of the Dutch Social Security Institute (SSI) between December 2016 and January 2017. All individuals were workers without a permanent employment contract, i.e. unemployed workers, temporary agency workers or workers with an expired fixed-term contract, sick-listed for at least two weeks. We excluded individuals who had been on sickness absence for less than two weeks as the probability to recover in this period is high, and therefore interventions for return to work are neither needed nor not cost-effective. In this study, we used a follow-up period of one year. We included all individuals still being sick-listed at the end of the one-year follow-up period and all individuals for whom the sickness absence benefit was ended because an individual had recovered. Individuals for whom the benefit was ended for other reasons, such as retirement, maternity leave or imprisonment, were excluded. The Medical Ethics Committee of Amsterdam UMC, VU University Medical Centre Amsterdam, gave ethical approval for this study and declared that no comprehensive ethical approval was needed.

In the Netherlands, sickness absence benefits for workers without an employment contract are assessed by the SSI. The SSI is a publicly funded agency that assesses benefit claims, takes care of benefit payments and provides re-integration support. Sickness absence benefits can be approved for a disease or handicap due to either social (i.e. non-occupational) or occupational causes and last for maximum two years. The SSI provides sickness absence benefits for all workers without a permanent employment contract (about 40% of the working population). In contrast, employers are responsible for continued payment of wages and re-integration support for their employees with a fixed contract. After two years of sickness absence, all individuals can apply for a disability benefit under the Dutch Work and Income Act (WIA).

Dependent variable

The dependent variable, long-term sickness absence, was based on sickness absence duration data as registered by the SSI and dichotomized: individuals who had long-term sickness absence (i.e. still being sick-listed at one-year follow-up) and individuals who did not have long-term sickness absence.

Independent variables

The aim of the prediction model was to identify, at the time of entry into the benefit, individuals who are at risk for sickness absence at one-year follow-up. Hence, all independent variables were measured at baseline. Part of the independent variables were retrieved from the databases of the SSI: the sociodemographics age, gender, marital status, and educational level, as well as the work-related characteristics work status and occupational sector, and number of sickness absence days in the past year. In addition, a number of work-related, health-related and psychosocial characteristics were collected by the SSI using self-reported questionnaires that individuals needed to fill out when applying for the sickness absence benefit. Answering the self-reported questionnaires was part of the SSI process and thus obligatory. Work-related variables included self-reports on return to work expectations and possibilities to apply for jobs (yes/no). From the Dutch National Questionnaire Working Conditions (NEA) the following questions were used about the last job before sickness absence: labour conflict, physically demanding job, mentally demanding job, and work demands. The response categories were dichotomous for all questions: "mostly physical" and "mostly mental" for the last question, and "yes" and "no" for all other questions [28].

Health-related variables included reason of sickness absence (categorized as "mental disorders", "musculoskeletal disorders", and "other physical disorders"), expected duration of sickness absence ("less than 1 month", "1-3 months", "more than 3 months", and "don't know"), and expected change in health during the next year ("improvement", "deterioration", and "no change"). General health condition was measured on a 5-point Likert scale ranging from "very bad" to "very good" [29]. Because only 37 individuals scored "very bad" on this question (<4% of the total study population), we merged the categories "bad" and "very bad".

Psychosocial factors were measured using the Well-Being Inventory (WBI) [30]. Individuals were asked whether they had problems with help-seeking, problem-solving, slowing down, ability to control events, whether they were worrying about the future in such extent that it prevented them from performing daily life activities, and whether they set high standards at work. The response options for all these variables were "yes" and "no".

Statistical analysis

Logistic regression analysis was used to determine prognostic factors to identify individuals with sickness absence at one-year follow-up. The model was build using three steps. First, we performed univariable analyses to test the association of each independent variable with the outcome variable using likelihood ratio (LR) tests (cut-off score $p>0.15$). Second, the variables remaining from the univariable analyses were tested for multicollinearity using variance inflation factors (VIFs). If VIF≥10, the strongest predictor for long-term sickness absence was chosen [31]. Third, we selected the subset of predictors for the final model using a hybrid approach combining forward and backward selection procedures, adhering to Akaike's Information Criterion (AIC) as stopping rule [32].

Calibration, i.e. the agreement between observed and predicted risk of sickness absence, of the prediction model was assessed using the Hosmer-Lemeshow goodness-of-fit test. A *p*-value ≥0.05 indicated that observed and predicted event rates were not significantly different. The discriminative ability of the model was evaluated using the area under the curve (AUC). The AUC is indicative of the percentage of correctly identified individuals at risk of long-term sickness absence. We interpreted AUC <0.60 as failing, 0.60-0.69 as poor, 0.70-0.79 as fair, 0.80-0.89 as good, and 0.90-1.00 as perfect discrimination [33].

In general, prediction models perform better in the sample used to fit the model than in an external sample. To obtain a more accurate estimate of model performance, the internal validity of the prediction model was examined by using a bootstrap approach [34]. We repeatedly drew 1,000 samples from the study cohort, with replacement, and calculated the corrected AUC by comparing the prediction model in the bootstrap samples with the original sample [35].

Latent class analysis was used to identify homogenous, mutually exclusive subgroups ("clusters") of sick-listed workers without an employment contract. Based on the predicted risk of sickness absence at one-year follow-up, we calculated tertiles and divided the individuals into three risk groups: individuals with a low, medium and high predicted probability of long-term sickness absence. The latent class analysis was based on the predicted risk groups and the independent variables.

Latent class analyses were conducted specifying two to five clusters. We used Bayesian Information Criterion (BIC) to assess model fit and determine the number of clusters in the optimal model [36, 37]. Individuals were assigned to the class with the highest posterior probability, i.e., to the class that best suited them. Average posterior class probabilities indicated the likelihood of class membership across all individuals whose maximum posterior probability was for that class and could be used to measure classification accuracy. The latent class analysis was considered accurate when the average posterior probabilities for all clusters were above 0.7 [36]. We interpreted the clusters based on the indicators with item-response probabilities of 0.7 or higher, as these indicators could be considered to be key characteristics of that cluster [38].

In general, all available variables can be used in latent class analysis. However, for practical purposes, selecting variables based on their usefulness for clustering was desirable as this improves interpretability of the model. Moreover, in the present study, most of the independent variables were retrieved from self-reported questionnaires, and shorter questionnaires are preferable in terms of costs and missing data. Therefore, we applied a variable selection approach based on the notion of BIC-based model selection [39]. Variables were sequentially considered for inclusion or exclusion from the set of variables selected for clustering based on their effect on BIC, maximized over the number of clusters and model parameterization.

All analyses were performed in RStudio for Windows, version 0.99.902.

Results

The study population contained 437 individuals. Table 1 shows the baseline characteristics of the study population. The median sickness absence duration was 105 (Interquartile range [IQR] 46-396) days. After one year, 133 individuals (30%) were still on sickness absence.

The final model predicting sickness absence at one-year follow-up included three variables as predictors: educational level, expected sickness absence duration, and help-seeking ability. Table 2 shows the coefficients of the final prediction model. The *p*-value of the Hosmer-Lemeshow goodness-of-fit test was 0.411, showing adequate calibration of the prediction model. The AUC of the final model was 0.777 (95% CI 0.731-0.822), showing fair discrimination for sickness absence at one-year follow-up. Using bootstrap validation, the optimism-corrected

AUC was 0.761 (95% CI 0.725–0.798). Multicollinearity was not assumed, as all VIF scores in the collinearity statistics for the multivariable model were <10.

The best fitting model in the latent class analysis was the model with four clusters based on seven variables. Table 3 presents the characteristics of the four clusters that were named: sick-listed workers with positive expectations, sick-listed workers with mental limitations, sick-listed workers with physical limitations, and sick-listed workers with negative expectations. The cluster of sick-listed workers with positive expectations consisted mainly of individuals with a good general health condition, but with temporary musculoskeletal or other physical disorders. The majority of these individuals expected to recover within three months and fully return to work afterwards. Generally, individuals in the mental limitations cluster had mild and temporary mental disorders. The majority had positive expectations about return to work, but they expected longer episodes of sickness absence than individuals in the positive expectations cluster. Sick-listed workers with physical limitations suffered mostly from musculoskeletal or other physical disorders with a longer recovery time. They expected their recovery to be within one month to more than three months. Individuals with more severe mental disorders made up the largest part of the cluster with negative expectations. They had a high risk of long-term sickness absence and negative coping skills.

There was a clear difference between the positive expectations cluster and the negative expectations cluster with respect to predicted risk of sickness absence and expected sickness absence duration: whereas all individuals in the positive expectations cluster expected to recover within 3 months, most individuals in the negative expectations cluster expected to be sick-listed for more than three months. Likewise, 67% of the individuals in the positive expectations cluster had a low risk of long-term sickness absence, while for 85% in the negative expectations cluster the model predicted a high risk. On the contrary, in the physical limitations cluster, both the expected sickness absence duration and the predicted risk of long-term sickness absence were much more varied. For sick-listed workers with negative expectations, the percentage with positive expectations about return to work was much lower (45%), than in the other three clusters. Concerning self-reported limitations and psychosocial factors, more than 75% of the individuals in the positive expectations and physical limitations clusters reported no difficulties with mental activities and positive coping skills. In the mental limitations and negative expectations clusters, the majority reported moderate to severe difficulties with mental activities and negative coping skills. The average posterior probabilities of the four clusters were 0.79, 0.88, 0.86 and 0.89, respectively, indicating good classification accuracy.

Table 1. Descriptive statistics of the study population at baseline

	Study population (n=437)	LTSA[a] (n=133)	Non-LTSA[a] (n=304)
Sociodemographics			
Age[b] (years)	44.9 [12.3]	45.3 [11.6]	44.7 [12.6]
Gender (female)	53%	62%	49%
Educational level[c]			
Low	34%	54%	25%
Secondary	39%	29%	43%
High	11%	14%	9%
Unknown	17%	4%	22%
Partner (yes)	65%	64%	66%
Work-related (characteristics of the previous job)			
Occupational sector			
Agriculture	11%	13%	11%
Finance	16%	17%	16%
Manufacturing	30%	26%	32%
Wholesale and retail	9%	11%	8%
Services	16%	22%	16%
Transportation	10%	7%	12%
Other	5%	5%	5%
Labour contract			
Unemployed workers	67%	73%	64%
Temporary agency workers	10%	7%	11%
Workers with an expired fixed-term contract	23%	20%	25%
Labour conflict (yes)	8%	12%	7%
Physically demanding job (yes)	59%	58%	60%
Mentally demanding job (yes)	46%	54%	43%
Work demands			
Mostly physical	63%	56%	65%
Mostly mental	37%	44%	35%
Return to work expectations (yes)	76%	67%	80%
Possibility to apply for jobs (yes)	34%	26%	38%
Health-related			
Reason of sickness absence			
Mental disorder	26%	32%	23%
Musculoskeletal disorder	40%	33%	43%
Other physical disorder	23%	18%	24%
Comorbidity of mental and physical disorders	12%	17%	10%
Number of sickness absence episodes in the past year[b]	0.24 [0.52]	0.24 [0.54]	0.24 [0.52]
Expected sickness absence duration			
Less than 1 month	12%	8%	13%
1-3 months	43%	28%	49%
More than 3 months	46%	65%	37%

3

Table 1. Continued

	Study population (n=437)	LTSA[a] (n=133)	Non-LTSA[a] (n=304)
General health condition			
(Very) bad	18%	23%	15%
Moderate	29%	35%	26%
Good	41%	36%	43%
Very good	13%	6%	16%
Expected health change			
No change	20%	23%	19%
Deterioration	8%	11%	7%
Improvement	72%	67%	74%
Limitations			
Difficulties with physical activities			
None	12%	6%	14%
Moderate	26%	27%	26%
Severe	62%	67%	60%
Difficulties with mental activities			
None	42%	33%	45%
Moderate	25%	23%	25%
Severe	34%	44%	29%
Relational or financial problems (yes)	24%	25%	24%
Psychosocial factors			
Help-seeking ability (yes)	55%	41%	61%
Worrying about the future (yes)	43%	50%	40%
Low control (yes)	53%	60%	50%
Problem-solving skills (yes)	64%	56%	67%
Set high standards at work (yes)	79%	84%	77%
Ability to slow down (yes)	27%	24%	29%

[a] LTSA = long-term sickness absence, i.e. individuals still receiving sickness absence benefit at one-year follow-up

[b] Average and standard deviation

[c] Based on the highest level of education completed. Low = primary school, lower vocational education, lower secondary school. Secondary = intermediate vocational education, upper secondary school. High = upper vocational education, university

Table 2. Coefficients of the final model predicting sickness absence at one-year follow-up

	OR [95% CI]	*p*-value
Educational level		
Low	1	
Secondary	0.34 [0.21-0.58]	0.000
High	0.67 [0.33-1.39]	0.283
Unknown	0.10 [0.04-0.26]	0.000
Expected sickness absence duration		
Less than 1 month	1	
1-3 months	1.17 [0.51-2.69]	0.712
More than 3 months	2.82 [1.26-6.39]	0.012
Help-seeking ability (yes)	0.59 [0.37-0.94]	0.027

Table 3. Characteristics of individuals in the four latent classes

	Positive expectations (n=82)	Mental limitations (n=105)	Physical limitations (n=138)	Negative expectations (n=112)
Risk of long-term sickness absence				
Low	67%	42%	31%	2%
Moderate	29%	47%	38%	13%
High	4%	11%	32%	85%
Return to work expectations (yes)	100%	86%	79%	45%
Reason of sickness absence				
Mental disorder	0%	56%	0%	48%
Musculoskeletal disorder	71%	5%	71%	13%
Other physical disorder	29%	21%	26%	13%
Comorbidity of mental and physical disorders	0%	18%	3%	26%
Expected sickness absence duration				
Less than 1 month	30%	14%	9%	0%
1-3 months	70%	49%	49%	10%
More than 3 months	0%	37%	43%	90%
Difficulties with mental activities				
None	78%	2%	75%	12%
Moderate	6%	51%	8%	70%
Severe	16%	47%	18%	18%
Help-seeking ability (yes)	85%	59%	69%	10%
Low control (yes)	15%	83%	28%	84%

3

Discussion

The aims of this study were to (i) predict sickness absence in a vulnerable group of workers without an employment contract at one-year follow-up, by building a model based on SSI registration data and self-reported questionnaires and (ii) explore whether distinct subgroups of sick-listed workers could be identified. The prediction model showed fair discrimination between individuals with and without long-term sickness absence based on three variables. Four types of sick-listed workers without an employment contract could be distinguished, partly based on the predicted risk of sickness absence at one-year follow-up.

The prediction model for sickness absence at one-year follow-up contained educational level, expected sickness absence duration, and help-seeking ability. The strongest predictor was self-reported expectations about sickness absence duration. This is in line with a previous study among sick-listed unemployed and temporary agency workers with psychological problems. That study reported that self-reported expectations about longer duration until full return to work was a strong prognostic factor for low work participation at long-term follow-up [17]. The other prognostic factor for long-term sickness absence in their final model was poor perceived health, which was not

found to be a predictor in the present study. This could be because in our model help-seeking ability was included, whereas their potential independent prognostic variables did not include psychosocial factors, or because their study population consisted only of workers with psychological problems, which could have influenced perceived health.

Whereas only a few have studied prognostic factors in workers without a permanent employment contract, several studies have focused on prognostic factors for sickness absence duration for sick-listed employees. These studies showed that also for employees there is a relation between self-reported expectations and sickness absence duration. Among a Dutch cohort of sick-listed teachers, expectation of duration of sickness absence longer than three months was found to be a predictor of longer time until return to work [15]. Likewise, other studies have found a significant relation between self-reported expectations and return to work for injured employees and employees on sick leave for at least four weeks [40, 41]. A relation between psychosocial factors and sickness absence duration has been demonstrated as well [42-44]. Lower educational level proved to be predictive of long-term sickness absence in a Swedish cohort of individuals on sick leave for at least 55 days [45]. As previous studies demonstrated these relations among cohorts of employees with an employment contract, we have shown that this relation also holds for sick-listed individuals without an employment contract.

We found an optimism-corrected AUC of 0.733 (95% CI 0.707–0.758) for the model predicting sickness absence at one-year follow-up. Previous studies on sickness absence duration for workers without a permanent employment contract did not report measures of the discriminative ability of the prediction models, thereby giving no information on the degree to which the predictions are valid for individuals from the underlying population [17]. Studies focusing on predicting sickness absence among employees did, and they found AUC values similar to our prediction model, i.e. ranging from 0.73 to 0.76 and showing fair discrimination between individuals with and without risk of long-term sickness absence [15, 23, 46]. However, most of these studies did not correct for over-optimism, and therefore their AUC values could be overestimated.

Four groups of sick-listed workers without an employment contract could be distinguished. Latent class analysis has previously been applied in occupational health studies concerning several populations, such as work disability for employees with diabetes and young adults with mental disorders [47, 48]. However, we are not aware of studies that applied latent class analysis to sick-listed employees without an employment contract. Latent class analysis is an effective method of data reduction and can guide stratified group-based intervention strategies. The results of the present study show that sick-listed workers in the negative expectations cluster, and possibly individuals in the physical limitations cluster, are most in need of return to work support as they have the highest risk of long-term sickness absence. Return to work interventions for these workers could be tailored at the characteristics of the clusters. For instance, workers in the negative expectations cluster are characterized by low self-control and being sick-listed due to (comorbidity of physical and) mental disorders. They might benefit from an intervention developed for sick-listed unemployed workers with psychological problems, like supported employment and interventions aiming at goal-setting and increasing the sense of control [49, 50]. Most workers in the physical limitations cluster are sick-listed due to musculoskeletal disorders. They are more likely to benefit from other types of interventions, such as a participatory return to work program or an intervention aimed at examination, information, and recommendations to remain active [51, 52]. Contrary, for individuals in the positive expectations and mental limitations clusters minimal support to return to work may be sufficient as they have a low risk of long-term sickness absence.

Strengths and limitations

A strength of the present study is the heterogeneous study population. We included all workers without an employment contract who were granted a sickness absence benefit by two regional offices of the SSI. In contrast to most previous studies on longer-term sickness absence that focused on individuals with specific diagnoses or individuals belonging to a certain occupational sector, our study population covered all diagnoses and occupational groups. As shown in previous studies, as well as in the prediction model of the present study, sickness absence duration is mostly determined by non-disorder related factors, and a prediction model that could be used for all diagnoses would be more useful for occupational health professionals. Second, as answering self-reported questionnaires was part of the SSI working process and obligatory for all individuals, there is no non-response and thus no selection bias. Moreover, our study population consisted of unemployed workers, temporary agency workers and workers with an expired fixed-term contract as these are the most vulnerable group within the working population. This means that our results are of interest for social security agencies and occupational health professionals. In addition, by using a variable selection algorithm for latent class analysis, we were able to find a parsimonious clustering. The clustering was partly based on self-reported questionnaires, and shorter questionnaires are preferred from a patient point of view. Moreover, as a parsimonious clustering is easier to interpret by occupational health professionals, it better suits practical needs.

Identifying subgroups of individuals based on statistical methods helps to obtain an unbiased classification, i.e. to reduce the influence of professionals' own values and judgements. However, a limitation of latent class analysis is that it could result in subgroups that are not recognizable by occupational health professionals. A combined approach of statistical methods and group consensus could be used to ensure a validated and practically relevant classification. Another limitation of the study is that the self-reported questionnaires were developed for practical purposes. Questions were selected based on considerations of professionals in the field of sickness absence services and a literature search. This means that the questionnaires used by the SSI consisted of a set of single questions from several (validated) questionnaires. Moreover, it is possible that not all relevant predictors were measured.

Practical implications

The longer individuals are absent from work, the less likely they are to return to work [1]. Therefore, it is important for policymakers and occupational health professionals to know which factors predict long-term sickness absence. The present study showed that only three variables might be needed to fairly discriminate between individuals with and without long-term sickness absence. As asking only a limited number of variables takes less time, it is preferred in terms of user-friendliness.

Some individuals are more vulnerable to long-term sickness absence than others, especially individuals with a low educational level, negative expectations of sickness absence duration and lower help-seeking ability. As individuals can be expected to make a good estimation of the duration of their sickness absence themselves based on experiences and personal and environmental factors, individuals who expect to recover in the short term may require less guidance from occupational health professionals than individuals with negative recovery expectations [53, 54].

Because occupational health services resources are limited, a differentiated approach is needed in occupational rehabilitation. Sick-listed workers without an employment contract are a heterogeneous group consisting of several more homogenous subgroups. Some subgroups might benefit more from return to work interventions than others. Hence, the latent class analysis results could be used as a triage tool to identify groups of claimants with a high risk of long-term sickness absence, get insight into the characteristics of these groups, and offer each group return to work interventions tailored to their characteristics. The results of the present study indicate that return to work interventions should at least be offered to individuals belonging to the negative expectations cluster, and, in case of sufficient capacity of occupational health services, probably also to individuals in the physical limitations cluster. On the other hand, individuals belonging to subgroups with a low risk of long-term sickness absence (i.e. sick-listed workers in the positive expectation and mental limitations clusters) are likely to recover themselves within one year without extensive support from occupational health professionals.

The predicted risk of long-term sickness absence and the partition of claimants into subgroups could be used by occupational health professionals at the start of the sickness absence period. It could be used as an additional source of information and guide professionals in selecting favourable return to work interventions for a particular claimant. During the rehabilitation process, new information might unfold and adjustment of the provided services might be needed. For instance, life events and differences in services or return to work interventions that sick listed workers receive might influence sickness absence duration. Hence, regular sickness absence monitoring is important to identify whether adjustment of return to work interventions might be beneficial.

Concluding remarks

This study showed that a logistic regression model could fairly discriminate between individuals with and without long-term sickness absence. Occupational health professionals could use the outcome of the prediction model to identify individuals at risk of long-term sickness absence. The allocation of workers into distinct groups could be used for efficient allocation of return to work interventions tailored to the groups that would most benefit from it.

References

1. Waddell G, Burton AK. Is work good for your health and well-being? London: The Stationery Office; 2006.
2. Schuring M, Robroek SJ, Otten FW, Arts CH, Burdorf A. The effect of ill health and socioeconomic status on labor force exit and re-employment: a prospective study with ten years follow-up in the Netherlands. Scand J Work Environ Health. 2013;39(2):134-143.
3. Henderson M, Harvey SB, Øverland S, Mykletun A, Hotopf M. Work and common psychiatric disorders. J R Soc Med. 2011;104(5):198-207.
4. Lännerström L, Wallman T, Holmström IK. Losing independence–the lived experience of being long-term sick-listed. BMC Public Health. 2013;13:745.
5. Louwerse I, Huysmans MA, van Rijssen HJ, van der Beek AJ, Anema JR. Characteristics of individuals receiving disability benefits in the Netherlands and predictors of leaving the disability benefit scheme: a retrospective cohort study with five-year follow-up. BMC Public Health. 2018;18(1):157.
6. Henderson M, Glozier N, Elliott KH. Long term sickness absence. BMJ. 2005;330:802.
7. Marhold C, Linton SJ, Melin L. A cognitive–behavioral return-to-work program: effects on pain patients with a history of long-term versus short-term sick leave. Pain. 2001;91(1-2):155-163.
8. van Aerden K, Puig-Barrachina V, Bosmans K, Vanroelen C. How does employment quality relate to health and job satisfaction in Europe? A typological approach. Soc Sci Med. 2016;158:132-140.
9. Herbig B, Dragano N, Angerer P. Health in the long-term unemployed. Dtsch Arztebl Int. 2013;88(3):363-368.
10. Virtanen M, Kivimäki M, Joensuu M, Virtanen P, Elovainio M, Vahtera J. Temporary employment and health: a review. Int J Epidemiol. 2005;34(3):610-622.
11. Benach J, Amable M, Muntaner C, Benavides FG. The consequences of flexible work for health: are we looking at the right place? 2002, BMJ Publishing Group Ltd.
12. Borrell-Carrió F, Suchman AL, Epstein RM. The biopsychosocial model 25 years later: principles, practice, and scientific inquiry. The Annals of Family Medicine. 2004;2(6):576-582.
13. De Witte H, Pienaar J, De Cuyper N. Review of 30 years of longitudinal studies on the association between job insecurity and health and well-being: is there causal evidence? Australian Psychologist. 2016;51(1):18-31.
14. Vermeulen SJ, Tamminga SJ, Schellart AJ, Ybema JF, Anema JR. Return-to-work of sick-listed workers without an employment contract–what works? BMC Public Health. 2009;9(1):232.
15. Nieuwenhuijsen K, Verbeek JH, de Boer AG, Blonk RW, van Dijk FJ. Predicting the duration of sickness absence for patients with common mental disorders in occupational health care. Scand J Work Environ Health. 2006;32(1):67-74.
16. Lammerts L, Schaafsma FG, Eikelenboom M, Vermeulen SJ, van Mechelen W, Anema JR, et al. Longitudinal associations between biopsychosocial factors and sustainable return to work of sick-listed workers with a depressive or anxiety disorder. J Occup Rehabil. 2016;26(1):70-79.
17. Audhoe SS, Hoving JL, Nieuwenhuijsen K, Friperson R, de Jong PR, Sluiter JK, et al. Prognostic factors for the work participation of sick-listed unemployed and temporary agency workers with psychological problems. J Occup Rehabil. 2012;22(4):437-446.
18. van Hoffen MF, Roelen CA, van Rhenen W, Schaufeli WB, Heymans MW, Twisk JW. Psychosocial work characteristics and long-term sickness absence due to mental disorders. J Ment Health. 2018; doi:10.1080/09638237.2018.1437603.

19. Cornelius L, Van der Klink J, Groothoff J, Brouwer S. Prognostic factors of long term disability due to mental disorders: a systematic review. J Occup Rehabil. 2011;21(2):259-274.
20. Holtermann A, Hansen JV, Burr H, Søgaard K. Prognostic factors for long-term sickness absence among employees with neck-shoulder and low-back pain. Scand J Work Environ Health. 2010;36(1): 34-41.
21. Lötters F, Burdorf A. Prognostic factors for duration of sickness absence due to musculoskeletal disorders. Clin J Pain. 2006;22(2):212-221.
22. Shiri R, Heliövaara M, Ahola K, Kaila-Kangas L, Haukka E, Kausto J, et al. A screening tool for the risk of disability retirement due to musculoskeletal disorders. Scand J Work Environ Health. 2018;44(1):37-46.
23. Bosman LC, Roelen CA, Twisk JW, Eekhout I, Heymans MW. Development of Prediction Models for Sick Leave Due to Musculoskeletal Disorders. J Occup Rehabil. 2019;29(3):617-624.
24. van Muijen P, Duijts SF, van der Beek AJ, Anema JR. Prognostic factors of work disability in sick-listed cancer survivors. J Cancer Surviv. 2013;7(4):582-591.
25. Roelen CA, Koopmans PC, Groothoff JW, van der Klink JJ, Bültmann U. Sickness absence and full return to work after cancer: 2-year follow-up of register data for different cancer sites. Psychooncology. 2011;20(9):1001-1006.
26. Demou E, Smith S, Bhaskar A, Mackay DF, Brown J, Hunt K, et al. Evaluating sickness absence duration by musculoskeletal and mental health issues: a retrospective cohort study of Scottish healthcare workers. BMJ open. 2018;8(1): e018085.
27. OECD. Sickness and Disability Schemes in the Netherlands: Country memo as a background paper for the OECD Disability Review. 2007.
28. Hooftman W, Mars G, Janssen B, De Vroome E, Ramaekers M, van den Bossche S. Nationale Enquête Arbeidsomstandigheden 2017. Methodologie en globale resultaten [National Working Conditions Survey 2017. Methodology and Overall Results]. TNO; 2018.
29. Ware Jr JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. Med Care 1992;30(6):473-483.
30. Vendrig A, Schaafsma F. Reliability and Validity of the Work and Well-Being Inventory (WBI) for Employees. J Occup Rehabil. 2018;28(2):377-390.
31. Kennedy P. A guide to econometrics. West Sussex: John Wiley & Sons; 2008.
32. Burnham KP, Anderson DR. Model selection and multimodel inference. New York: Springer; 2002.
33. Obuchowski NA. Receiver operating characteristic curves and their use in radiology. Radiology. 2003;229(1):3-8.
34. Steyerberg EW, Harrell Jr FE, Borsboom GJ, Eijkemans M, Vergouwe Y, Habbema JDF. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54(8):774-781.
35. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361-387.
36. Masyn KE. 25 latent class analysis and finite mixture modeling. The Oxford handbook of quantitative methods. 2013.
37. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Struct Equ Modeling. 2007;14(4):535-569.
38. Collins LM, Lanza ST. Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences. New Jersey: John Wiley & Sons; 2010.
39. Dean N, Raftery AE. Latent class analysis variable selection. Ann Inst Stat Math. 2010;62(1):11.

3

40. Cole DC, Mondloch MV, Hogg-Johnson S, et al. Listening to injured workers: how recovery expectations predict outcomes—a prospective study. CMAJ. 2002;166(6):749-754.
41. Volker D, Zijlstra-Vlasveld M, Brouwers E, Van Lomwel A, Van der Feltz-Cornelis C. Return-to-work self-efficacy and actual return to work among long-term sick-listed employees. J Occup Rehabil. 2015;25(2):423-431.
42. Brouwer S, Krol B, Reneman MF, Bültmann U, Franche R-L, van der Klink JJ, et al. Behavioral determinants as predictors of return to work after long-term sickness absence: an application of the theory of planned behavior. J Occup Rehabil. 2009;19(2):166-174.
43. Post M, Krol B, Groothoff JW. Self-rated health as a predictor of return to work among employees on long-term sickness absence. Disabil Rehabil. 2006;28(5):289-297.
44. Lund T, Christensen KB, Vaez M, Labriola M, Josephson M, Villadsen E, et al. Differences in sickness absence in Sweden and Denmark: the cross national HAKNAK study. Eur J Public Health. 2008;19(3):343-349.
45. Karlsson NE, Carstensen JM, Gjesdal S, Alexanderson KA. Risk factors for disability pension in a population-based cohort of men and women on long-term sick leave in Sweden. Eur J Public Health. 2008;18(3):224-231.
46. Roelen CA, Bültmann U, van Rhenen W, van der Klink JJ, Twisk JW, Heymans MW. External validation of two prediction models identifying employees at risk of high sickness absence: cohort study with 1-year follow-up. BMC Public Health. 2013;13(1):105.
47. Virtanen M, Vahtera J, Head J, Dray-Spira R, Okuloff A, Tabak AG, et al. Work disability among employees with diabetes: latent class analysis of risk factors in three prospective cohort studies. PloS one. 2015;10(11).
48. Joensuu M, Mattila-Holappa P, Ahola K, Ervasti J, Kivimäki M, Kivekäs T, et al. Clustering of adversity in young adults on disability pension due to mental disorders: a latent class analysis. Soc Psychiatry Psychiatr Epidemiol. 2016;51(2):281-287.
49. Audhoe SS, Nieuwenhuijsen K, Hoving JL, Sluiter JK, Frings-Dresen MH. The effectiveness of the "Brainwork Intervention" in reducing sick leave for unemployed workers with psychological problems: design of a controlled clinical trial. BMC public health. 2015;15(1):377.
50. Suijkerbuijk YB, Schaafsma FG, van Mechelen JC, Ojajärvi A, Corbière M, Anema JR. Interventions for obtaining and maintaining employment in adults with severe mental illness, a network meta-analysis. Cochrane Database of Systematic Reviews. 2017(9).
51. Vermeulen SJ, Anema JR, Schellart AJ, Knol DL, van Mechelen W, van der Beek AJ. A participatory return-to-work intervention for temporary agency workers and unemployed workers sick-listed due to musculoskeletal disorders: results of a randomized controlled trial. J Occup Rehabil. 2011;21(3):313-324.
52. Hagen EM, Eriksen HR, Ursin H. Does early intervention with a light mobilization program reduce long-term sick leave for low back pain? Spine. 2000;25(15):1973-1976.
53. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev. 1977;84(2):191-215.
54. Flood AB, Lorence DP, Ding J, McPherson K, Black NA. The role of expectations in patients' reports of post-operative outcomes and improvement following therapy. Med Care 1993;31(11):1043-105.

3

CHAPTER 4

Predicting future changes in work ability of individuals receiving a work disability benefit: weighted analysis of longitudinal data

I Louwerse
MA Huysmans
HJ van Rijssen
FG Schaafsma
KHN Weerdesteijn
AJ van der Beek
JR Anema

Published in:
Scandinavian Journal of Work, Environment and Health
2020;46(2):168-176

Abstract

Background Weighted regression procedures can be an efficient solution for cohort studies that involve rare events or diseases, which can be difficult to predict, allowing for more accurate prediction of cases of interest. The aims of this study were to (i) predict changes in work ability at one year after approval of the work disability benefit and (ii) explore whether weighted regression procedures could improve the accuracy of predicting claimants with the highest probability of experiencing a relevant change in work ability.

Methods The study population consisted of 944 individuals who were granted a work disability benefit. Self-reported questionnaire data measured at baseline were linked with administrative data from Dutch Social Security Institute databases. Standard and weighted multinomial logit models were fitted to predict changes in the work ability score (WAS) at one-year follow-up. McNemar's test was used to assess the difference between these models.

Results A total of 208 (22%) claimants experienced an improvement in WAS. The standard multinomial logit model predicted a relevant improvement in WAS for only 9% of the claimants [positive predictive value (PPV) 62%]. The weighted model predicted significantly more cases, 14% (PPV 63%). Predictive variables were several physical and mental functioning factors, work status, wage loss, and WAS at baseline.

Conclusion This study showed that there are indications that weighted regression procedures can correctly identify more individuals who experience a relevant change in WAS compared to standard multinomial logit models. Our findings suggest that weighted analysis could be an effective method in epidemiology when predicting rare events or diseases.

Introduction

Long-term work disability is bad for an individual's health, and returning to work is generally associated with a positive effect on the future course of the disease and work ability [1-3]. Individuals who are unable to work due to a disease or disorder can apply for a work disability benefit. In most European countries, this covers both financial support to compensate loss of income and interventions supporting return to work.

Possible predictors for work disability include a broad range of external and personal factors. When conducting medical disability assessments to evaluate whether a work disability benefit should be granted, insurance physicians (IPs) predominantly rely on factors relating to the disease and the disorder of a claimant [4, 5]. One of the main tasks of an IP is estimating prognosis of work disability and determining if and when a re-assessment should be planned [6]. Medical re-assessments are conducted to determine whether an individual's health has improved or deteriorated to such an extent that adjustment of support to return to work is required or the continuing eligibility for the benefit has changed. In The Netherlands, claim duration for work disability benefits is long lasting for most claimants and IPs consider prognosis of work disability as the most difficult part of the work disability assessment [7, 8]. Therefore, accurate prognosis of future changes in work disability is important to identify those in need of return to work interventions and for efficient planning of medical re-assessments.

Work ability, commonly measured with the Work Ability Index (WAI), is an important concept in the context of work disability duration. It is defined as the physical, mental, and social fit of an individual with the work demands and capability to participate in work [9]. Self-assessed work ability is a strong predictor of work disability duration and return to work [10, 11]. Clinical decision-support systems, in which characteristics of individual patients are used to generate patient-specific assessments or recommendations that are then presented to clinicians for consideration, are designed to aid decision-making [12]. They can optimize the time with the client and improve the overall quality of services [13]. A prediction model for future changes in work ability could aid IPs in their medical disability assessment and lead to more precise estimation of future work disability. Because resources to perform medical re-assessments are limited, the model is of most added value in practice if it can sufficiently identify claimants who will improve in their work ability. This ensures that medical re-assessments are planned at the time an assessment interview with an IP has the most added value. However, claimants who perceive a relevant future improvement of their work ability form only a relatively small proportion of the total number of work disability claimants.

Predicting rare events or diseases with probabilistic statistical regression is difficult, as these methods tend to be biased towards the majority class and underestimate the probability of rare events [14]. Weighted regression can take account of the preponderance of claimants not experiencing a substantial change in their work ability, and focus accuracy on claimants who most likely will experience a change. Weighted least squares have its origin in econometrics and are used in a range of application areas, such as psychology, regional science and time series analysis [15, 16]. However, we are not aware of any research in occupational epidemiology using weighted analysis. Therefore, the aim of this study was twofold: to (i) predict changes in work ability of claimants at one year after approval of the work disability benefit by building a model based on sociodemographic, work disability, health, functional limitation and personal factors; and (ii) explore whether the accuracy of predicting claimants with the highest probability of experiencing a relevant change of work ability could be improved by using weighted regression.

Methods

Study population

We used data of the FORWARD study, a longitudinal cohort study among 2539 individuals who applied for a work disability benefit at the Dutch Social Security Institute (SSI) between July 2014 and March 2015, after a two-year period of sick leave. Individuals were aged 18–64 years at inclusion. Claimants suffering from severe mental, cognitive, or visual disorders or those diagnosed with cancer were excluded from the FORWARD study. A more extensive description of the study cohort can be found elsewhere [17].

From the FORWARD study, we retrieved data from the baseline questionnaire completed just before the medical disability assessment and the questionnaire at one-year follow-up. For each participant, we combined the self-reported data of the cohort study with administrative data from SSI databases. The participants of the FORWARD study all signed informed consent. The Medical Ethics Committee of the VU University Medical Center (Amsterdam, The Netherlands) has approved the FORWARD study.

Inclusion and exclusion criteria

To be included in the present study, the single-item question of the WAI needed to be answered both at baseline and one-year follow-up. Of the 2,593 individuals included in the FORWARD study, 42 and 646 participants were excluded because they did not answer this question at baseline and one-year follow-up, respectively. We excluded participants who were ineligible or did not apply for work disability benefits (n=701) and those who were granted a permanent work disability benefit (n=260). In the latter case, there are no possibilities to return to work, and hence no re-assessments need to be scheduled. In total, 944 participants were included in the present study.

Dependent variable

The dependent variable of the model was the change in self-reported work ability at one-year follow-up as compared to baseline. Work ability was measured with the first question of the WAI, also referred to as the work ability score (WAS) [18]. This question asks participants to compare their current work ability with their lifetime best on a 0–10 scale. Higher scores indicate better work ability. The WAS is significantly correlated to the WAI and can therefore be used as a simple indicator for assessing work ability [19, 20]. A single-item measure takes less time to complete and analyse, and is therefore preferable in terms of costs, interpretation and missing data.

In line with previous studies we defined an improvement or deterioration in WAS of ≥2 points as the smallest detectable self-reported change likely to have an effect on job opportunities and work disability benefit [21, 22]. Based on their change in WAS scores at one-year follow-up as compared to baseline, we divided the participants into three groups: participants with no relevant change ($|WAS_{T1}-WAS_{T0}|\leq 1$), an improvement ($WAS_{T1}-WAS_{T0}\geq 2$), or a deterioration ($WAS_{T1}-WAS_{T0}\leq -2$), with WAS_{T0} and WAS_{T1} the scores at baseline and one-year follow-up. WAS_{T0} was also added as an independent variable to the model.

Independent variables

All independent variables were measured at baseline. The sociodemographics age, gender, marital status, and educational level, as well as the work-related characteristics work status and occupational sector were retrieved from the SSI database. In addition, a number of health characteristics were determined: primary diagnosis, comorbidity, permanency, treatment and medication, and functional limitations as registered by the IP during the medical disability assessment in the list of functional abilities (LFA). The LFA is partly based on the World Health Organisation's International Classification of Functioning, Disability and Health (ICF) [23]. It consists of 106 items indicating the presence (dichotomous) and severity (ordinal) of limitations, categorized into six sections: personal functioning, social functioning, adjusting to the physical environment, dynamic movements, static posture, and working hours. Higher scores on the ordinal rating scales indicate more severe limitations to perform activities. We considered the average number of limitations of the first five sections and the single question of the last section regarding restrictions in the working hours per day as independent variables. If a claimant is too seriously disabled to return to work, e.g., bedridden or receiving inpatient care, limitations are not registered in the LFA. This was the case for 119 (13%) of the participants in our study sample.

Besides registration data from the SSI, a number of self-reported surveys from the FORWARD study baseline questionnaire was used. The Short Form Health Survey (SF-36) is a measure of health status, containing 36 items on physical and mental functioning and role limitations, well-being, pain, general health, and health change. Scores range between 1–60, higher scores indicating better health status [24]. The Whitely Index (WI) contains 14 items to measure health anxiety. Scores range between 0–14, with higher scores indicating more severe health anxiety [25]. The Hospital Anxiety and Depression Scale (HADS) produces scales for anxiety and depression. Scores range between 0-21, higher scores indicating higher distress [26]. The Work and Wellbeing Inventory (WBI) measures symptoms, coping, support, stress and disability with 87 items. Scores range between 0-84, higher scores indicating more barriers for return to work [27]. We also retrieved household characteristic, and work-related characteristics regarding work demands and managerial tasks. The questionnaire also asked respondents about their expectations with respect to recovering and getting back to work.

Statistical analysis

Multinomial logistic regression analysis was used to predict changes in work disability at one-year follow-up. We fitted both standard and non-parametric multinomial logit (MNL) estimates. See figure 1 for the specification of the non-parametric MNL estimates. Because we were most interested in accurately predicting the largest improvements in WAS, we used the following linear weight function for claimants who experience an improvement in WAS (i.e. WAS_{T1}-$WAS_{T0} \geq 2$):

$$w_i = \tfrac{1}{2}(WAS_{T1} - WAS_{T0}) + 1 \qquad (1)$$

For all claimants who did not experience an improvement in WAS (i.e. WAS_{T1}-$WAS_{T0}<2$), the weight was set to $w_i = 1$. By using (1), claimants with an improvement in WAS of 2 points were assigned twice as much weight as claimants not experiencing an improvement in WAS. Because larger weights were assigned to claimants with a larger improvement in WAS, the model focusses on accurately predicting these claimants. In application areas where weighted regression is more often used, weight functions are often linear or exponential functions. For instance, in geographically weighted regression, locations that are closer get higher weights. In time series analysis, weights decrease for observations further back in time. Hence, the linear weight function (1) is in line with weight specifications in other research [15, 28]. Because weighted regression procedures are not commonly used in occupational epidemiology, there is no general approach to specify the exact weights that should be given to observations. Hence, we tried several weight functions and examined the effect on the performance of the prediction model. Assigning a weight equal to one to claimants with an improvement of WAS would result in the standard MNL model. Therefore, we considered assigning weights equal to 1.5, 2, 2.5, or 3. We did not consider weights >3 as we felt this would place disproportional emphasis on claimants with an improvement in WAS. Although the differences between the weight functions were small, we chose a weight of 2 in the final model as this resulted in the highest sensitivity, i.e., the model that could identify most claimants with an improvement in WAS. The

Let $j = 1,2,3$ denote the alternative categories that a claimant can belong to, based on the change in work ability at one-year follow-up, and let $i = 1, \dots, n$ denote the claimants. The probabilities p_{ij} of claimant i belonging to category j of the multinomial logit model are

$$p_{ij} = \text{Prob}[Y_i = j|x_i] = \frac{\exp(x_i'\beta_j)}{1+\sum_{k=2}^{J}\exp(x_i'\beta_k)} \qquad (2)$$

where x_i represents the characteristics of claimant i, and β_j measures the relative weights of the characteristics. The multinomial logit model can be estimated by maximum likelihood, i.e., by maximizing the log-likelihood

$$\log(\text{L}) = \sum_{i=1}^{n}\sum_{j=1}^{3} I_{ij}\log(p_{ij}) \qquad (3)$$

with respect to the parameters β_j, $j = 1,2,3$. Here, I_{ij} is an indicator variable, with $I_{ij} = 1$ if $Y_i = j$ and $I_{ij} = 0$ otherwise.

Now, let w_i be the weight given to claimant i. Instead of minimizing (2) we could minimize the following pseudo log-likelihood function

$$\log(\text{L}) = \sum_{i=1}^{n} w_i\left[\sum_{j=1}^{3} I_{ij}\log(p_{ij})\right]. \qquad (4)$$

Note that (4) includes standard multinomial logit as the special case with weights $w_i = 1$ for all i.

Figure 1. Specification of the non-parametric MNL estimates

positive predictive value (PPV) and negative predictive value (NPV) were similar for the different weight functions that were considered.

The models were built using three steps. First, we performed univariable analyses to test the association of each independent variable with the outcome variable using likelihood ratio (LR) tests (cut off score *p*>0.2). Second, the variables remaining from the univariable analyses were tested for multicollinearity using variance inflation factors (VIF). We considered VIF <10 to be acceptable [29]. Third, we selected the subset of predictors for the final model using a hybrid approach combining forward and backward selection procedures.

Before the start of the analysis, we randomly split the data into a training set (80% of the study population) to fit the models and a test set (20% of the study population) to evaluate the models. The purpose of developing the prediction model is that it can be used in practice. This means that we want to know how well the model predicts new cases. Therefore, the test set, i.e., the held-out sample, is used to get an unbiased estimate of model effectiveness.

We calculated several performance measures to compare the standard and weighted MNL model. We reported both specificity and sensitivity as these are important measures of diagnostic accuracy of a model. However, they are of no practical use when IPs need to estimate the probability of improvement in WAS for individual claimants [30]. Hence, predictive values are more meaningful performance measures in this context. In general, there is a trade-off between sensitivity and predictive values. We can indicate the added value of the weighted model if it results in predictions with both higher sensitivity and predictive values.

We used McNemar's test to statistically assess whether the standard and the weighted model had a similar proportion of errors on the test set. Calculation of the test statistic is based on the contingency table. It tests whether the models have equal accuracy for predicting true improvements in WAS, i.e., it detects whether the difference between the misclassification rates of the models is statistically significant. The level of significance was set at $p<0.05$. All analyses were performed in RStudio for Windows, version 0.99.902.

Results

Tables 1 and 2 show the baseline characteristics of the study population. Mean WAS on baseline was 2.5 [standard deviation (SD) 2.1], and 2.8 (SD 2.2) at one-year follow-up. The majority of the study population (n=599; 63%) did not experience a change in WAS at one-year follow-up; 208 claimants (22%) experienced an improvement in WAS (mean WAS improvement 3.1; SD 1.5) and 127 a deterioration (15%).

In this section, we mainly focus on the results of the 187 claimants who were randomly selected to be included in the test set. Among this group, the percentage experiencing a WAS improvement at one-year follow-up was slightly higher than that of the training set (24% versus 21%). Of all cases in the test set, the standard model predicted for 16.9% of the total number of claimants an improvement of the WAS at one-year follow-up (table 3). The sensitivity was only 22%, showing that it was difficult to identify relevant claimants with standard regression procedures. The PPV was 62% and the NPV 79%. Eight variables ended up in the standard model: WAS at baseline, work status, WBI disability, wage loss, SF36 energy, SF36 physical functioning, WBI symptoms, and WI.

The weighted model predicted a larger number of improvements compared to the standard model (table 3). The number of predicted cases increased from 16 to 27, i.e., from 9% to 14% of the total number of claimants, and was now closer to the percentage of actually observed improvements in the study population (22%). The PPV and NPV were 63% and 82%, respectively. The weighted model contained 11 variables. It included the same variables as the standard model, except for the variable WI. Additionally, the variables LFA static posture, LFA working hours, mental healthcare, and SF36 health change were added. All the VIF scores in the collinearity statistics for the multivariable models were <10, therefore multicollinearity was not assumed. The last two columns of table 1 show the coefficients of the multivariable logit models.

Table 1. Descriptive statistics of the study population at baseline representing number (n), percentage (%), mean and standard deviation (SD), and coefficients and 95% confidence intervals of the variables included in the multivariable logit models

	Study population (n=944)				Standard MNL[a] model	Weighted MNL[a] model
	n	%	Mean	SD	Coeff [95% CI]	Coeff [95% CI]
Occupational						
Work status (working)	200	21			1.03 [0.54,1.52]	1.03 [0.64,1.42]
Health						
Mental healthcare (yes)	487	52				-0.18 [-0.51,0.17]
Disability assessment						
Wage loss (≥80%)	548	58				-0.61 [-0.98,-0.25]
LFA[b] static posture			0.33	0.22	-0.64 [-1.12,-0.17]	-1.39 [-2.22,-0.55]
LFA[b] working hours per day						
>8 hours	398	42				1
≤8 hours	87	9				-0.05 [-0.59,0.49]
≤6 hours	96	10				-0.22 [-0.77,0.33]
≤4 hours	204	22				-0.08 [-0.12,0.04]
≤2 hours	40	4				-0.11 [-0.19,-0.02]
Unknown	119	13				-0.04 [-0.73,0.64]
Self-reported surveys						
SF36[c]						
Physical functioning			41.6	24.8	0.01 [0.00,0.02]	0.01 [-0.01,0.00]
Energy			30.5	17.6	0.02 [0.00,0.03]	0.02 [0.01,0.03]
Health change			37.5	27.8		0.01 [0.00,0.01]
Whitely Index			6.1	3.0	-0.03 [-0.10,0.05]	
Well-being inventory						
Symptoms			48.4	13.0	0.02 [-0.01, 0.04]	0.01 [-0.01,0.02]
Disability			24.1	4.2	-0.06 [-0.11,-0.01]	-0.05 [-0.10,-0.01]
Work ability score			2.5	2.1	-0.47 [-0.61,-0.33]	-0.55 [-0.66,-0.44]

[a] MNL = multinomial logit model

[b] LFA = list of functional abilities

[c] SF36 = Short Form Health Survey, 36 items

Table 2. Descriptive statistics of the study population at baseline of the variables not included in the multivariable models

	Study population (n=944)			
	n	%	Mean	SD
Sociodemographics				
Age (years)			51.2	9.0
Gender (female)	476	50		
Educational level				
Low	309	33		
Secondary	399	39		
High	266	28		
Partner (yes)	705	75		
Children (yes)	704	75		
Principal wage earner (yes)	629	67		
Occupational				
Occupational sector				
Finance	127	13		
Government	104	11		
Healthcare	204	22		
Manufacturing	104	11		
Wholesale and retail	120	13		
Other	285	30		
Managerial tasks (yes)	216	23		
Work demands				
Physical	271	29		
Psychological	285	30		
Physical and psychological	388	41		
Health				
Primary diagnosis				
Cardiovascular	96	10		
Mental	233	25		
Musculoskeletal	373	40		
Nervous system	87	9		
Other	155	16		
Comorbidity (yes)	669	71		
Medication use	840	89		
Disability assessment				
Possibility to work (yes)	789	84		
Permanency (yes)	304	32		
LFA[a] personal functioning			0.08	0.07
LFA[a] social functioning			0.11	0.12
LFA[a] adjusting to physical environment			0.11	0.09
LFA[a] dynamic movement			0.26	0.14

Table 2. Continued

	Study population (n=944)			
	n	%	Mean	SD
Self-reported surveys				
SF36[b]			6.8	19.9
Role limitations due to emotional problems			33.7	44.2
Emotional well-being			50.3	22.6
Social functioning			53.6	10.4
Pain			37.8	24.9
General health			33.4	17.0
HADS[c]				
Anxiety			9.5	4.8
Depression			9.8	5.0
Well-being inventory				
Coping			42.5	10.0
Support			56.4	12.3
Stress			37.9	9.5

[a] LFA = list of functional abilities
[b] SF36 = Short Form Health Survey, 36 items
[c] HADS = Hospital Anxiety and Depression Scale

Table 3. Predictions of the standard and weighted model (test set)

			Observed		
			Deterioration	No change	Improvement
			n (%)	n (%)	n (%)
Predicted	Standard model	Deterioration	7 (70)	3 (30)	0 (0)
		No change	25 (16)	100 (62)	36 (22)
		Improvement	0 (0)	6 (38)	10 (62)
	Weighted model	Deterioration	8 (57)	6 (43)	0 (0)
		No change	24 (16)	93 (64)	29 (20)
		Improvement	0 (0)	10 (37)	17 (63)

Table 4. Performance measures of the models

	MNL[a] model	
	Standard (%)	Weighted (%)
Sensitivity	22	37
Specificity	96	93
PPV[b]	62	63
NPV[c]	79	82

[a] MNL = multinomial logit model
[b] PPV = positive predictive value
[c] NPV = negative predictive value

4

The sensitivity, i.e., the model's ability to correctly detect claimants with an improvement in the WAS, increased from 22% to 37% when we compared the weighted to the standard model (table 4). Both the PPV and NPV of the weighted model were slightly higher as well; the PPV increased from 62% to 63%, and the NPV increased from 79% to 82%. This means that the predictions of the weighted model were correct more often than the predictions of the standard model, although the differences were small.

McNemar's χ^2 was equal to 6.667 and a corresponding *p*-value of 0.0009. This means that the two models had a different proportion of errors on the test set. The contingency table showed that the number of cases that the weighted model predicted correctly was higher than the number of claimants correctly classified by the standard model. The total number of claimants who were classified differently by the weighted model compared to the standard model was 15, which was sufficiently large to provide accurate *p*-values for McNemar's test (minimum number is 10) [31].

The results that the weighted model was better at predicting claimants who will experience an improvement in WAS at one-year follow-up for the test set were in line with the results of the training set. In the test, set the percentage of claimants identified increased from 9% to 14%.

Discussion

The aims of this study were to (i) predict changes in work ability at one year after approval of the work disability benefit and (ii) explore whether weighted regression procedures could improve the accuracy of predicting claimants with the highest probability of experiencing an improvement in WAS. A minority of 22% of the claimants in our study population experienced an improvement in WAS. Our standard model predicted a relevant improvement in WAS for only 9% of the claimants, while the weighted model predicted this for 14%. However, the PPV of the weighted model did not decrease compared to the standard model. Likewise, the NPV slightly increased. Hence, the weighted model predicted more claimants who will experience a relevant improvement in WAS at one-year follow-up. At the same time, IP can be more certain that the model predicts the correct outcome.

We used a weighted regression model with a linear weight function that assigns larger weights to claimants with a bigger improvement in WAS. Our finding that the weighted model could correctly identify a larger group of individuals with an improvement in WAS in both the training and test sets implies that our weight function could also be of added value in a population that was not used to build the models. However, as the set of possible weight functions is inexhaustible, it could be that there are other weight functions that provide similar or better results than the weight function we have chosen.

The majority of individuals in the study population (63%) did not experience a change in WAS at one-year follow-up. This is in agreement with previous research showing that changes in WAI are small for most individuals, especially for those with longer episodes of sickness absence [19, 32]. Determinants of work ability have been reported in several studies. In the present study, work ability at baseline was the strongest predictor in both models. This is in line with previous research showing that, for sick listed workers diagnosed with cancer, WAS at baseline was an important predictor for WAS at one-year follow-up [33]. This study also showed an association with wage loss, as we found that individuals with a lower level of wage loss were more likely to experience an improvement in WAS. A higher level of wage loss means more extensive functional limitations, which seems to have a negative effect on work ability at one-year follow-up. This relation was also found for degree of sickness absence and changed WAS at 6- and 12-months follow-up for women on sick leave for ≥60 days [19]. Several studies have also found a relation between the WAI and mental and physical conditions, demands at work, individual characteristics and lifestyle [34, 35]. These studies did, however, not report measures of diagnostic accuracy (e.g. sensitivity and predictive values) of the estimated models.

As pointed out in a recent editorial on prediction models for sickness absence, researchers should be careful making claims on accuracy of these models [36]. Although the difference between the standard and weighted model in terms of predicting claimants with an improvement in WAS statistically significant, it was small and it is therefore questionable whether this difference is relevant. However, in the current policies of the SSI, because of the limited capacity to perform IP re-assessments and the fact that only a minority of 22% of the individuals actually experienced an improvement in WAS at one-year follow-up, the prediction model may be a relevant tool for identifying the group of claimants with the highest probability of experiencing an improvement in WAS. Our focus was not on predictions at the individual level, but at a population level. Hence, the small differences between the standard and the weighted model are regarded as useful in achieving a more effective allocation of limited occupational health resources. The weighted model identifying 14% of the claimants, as opposed to 9% with the standard model, with 63% accuracy is considered as a useful auxiliary tool for IPs when they plan re-assessments. Likewise, in case the model predicts no substantial improvement in WAS at one-year follow-up (which is the case for 86% of the claimants), this could be an indication that for this group of claimants scheduling a re-assessment at one-year follow-up has less added value as the NPV is 82%. These probabilities are much higher than the case where the SSI policy is to plan re-assessments at random. However, it could be argued that, in other applications, the differences between the two models shown in the present study are too small to be of practical relevance.

We are not aware of any prediction model for future changes in work ability for individuals with a work disability benefit. Previous studies on long-term sickness absence in the general working population have shown that it is difficult to develop prediction models with high prediction accuracy that are relevant in practice. Studies identifying claimants at risk for work disability and long-term sickness absence showed only moderate prediction accuracy [37, 38]. Studies on prediction models for individuals with specific chronic diseases such as low back pain or common mental disorders validated prediction models in terms of PPV and NPV [39-41]. Similar to the results of the present study, the NPV of their models were in the range of 74-98%, which is considered high. However, they reported PPV of 33-57%, which is lower than the PPV of our model (63%).

Strengths and limitations

A strength of the present study is that, by fitting weighted MNL, we are better able to meet practical needs. Non-parametric models offer important advantages because they can focus accuracy on claimants who most likely will experience a change in their entitlement of the work disability benefit. Moreover, by dividing the study sample in a training set to build our prediction models on and a test set to validate the models, we were able to assess the predictive accuracy and generalization of the model. A further strength is that we combined self-reported questionnaire data with administrative data. This enriches the understanding of a broad range of medical, social, psychological, and work-related factors that can influence future work ability.

Moreover, whereas most studies about predictors of work disability duration and return to work focus on a specific category of diagnoses, our study cohort included a broad range of diseases and disorders. A limitation of our study is that two groups of individuals were excluded from the FORWARD cohort and could therefore not be included in our study: individuals suffering from severe mental, cognitive, or visual disorders (e.g., dementia or psychosis), due to their reduced ability to correctly complete the questionnaires, and individuals diagnosed with cancer.

A study limitation is that the FORWARD cohort questionnaires were not designed to identify the best independent variables for predicting changes in work ability. For instance, own expectation about future changes in work ability were not covered in the questionnaire while the individual's own expectations are important predictors for duration of long-term sick leave and return to work [42, 43]. Moreover, the administrative data that we used was not collected for research purposes but rather registered by SSI employees for administration purposes. However, the FORWARD cohort questionnaires are extensive and, by combining them with administrative data, we were able to cover a broad range of potential predictors. A final limitation of this study is our reliance on changes in self-reported work ability. In line with previous studies, we defined an improvement or deterioration in WAS of ≥2 points as a relevant change [10, 11, 21]. However, it should be investigated if this is also the case for our study population.

Implications for research and practice

Commonly reported outcomes in epidemiological and medical research, such as the incidence of clinical events among a cohort of patients or the response rate in patients taking a certain treatment regimen, are rare events and usually difficult to estimate. Disease predictions can contribute to a wide range of applications, such as risk management, tailored health communication, and decision support systems [44, 45]. Weighted analysis could aid these applications by making more accurate predictions of rare events and diseases.

Identification of claimants with a high probability of experiencing an improvement of work ability at one-year follow-up may assist IP during the medical disability assessment when they need to predict future work ability. This can aid accurate prognosis of work ability and providing suitable interventions to return to work.

To be used in practice, the prediction model needs to be supported by a suitable tool, which is easy to access and interpret for professionals. Future research should focus on the preferable design and content of such a decision support tool. Next, a cost-effectiveness analysis and process evaluation should be performed to determine the added value of the model for IP in making accurate prognoses of work ability.

Concluding remarks

This study showed that, compared to standard MNL models, there are indications that weighted regression procedures can correctly identify more claimants who experience an improvement in WAS. Our findings suggest that a weighted analysis could be an effective method in epidemiology when predicting rare events or diseases. More research is needed to examine the added value of weighted regression procedures in occupational epidemiology.

References

1. Waddell G, Burton AK. Is work good for your health and well-being? 2006, London: The Stationary Office.
2. Viikari-Juntura E, Kausto J, Shiri R, Kaila-Kangas L, Takala EP, Karppinen E, et al. Return to work after early part-time sick leave due to musculoskeletal disorders: a randomized controlled trial. Scand J Work Environ Health. 2012;38(2):134-143.
3. OECD. Sickness, Disability and Work: Breaking the Barriers: A Synthesis of Findings across OECD Countries. 2010, Paris: OECD Publishing.
4. Krause N, Frank JW, Dasinger LK, Sullivan TJ, Sinclair SJ. Determinants of duration of disability and return-to-work after work-related injury and illness: Challenges for future research. Am J Ind Med.2001;40(4):464-484.
5. Slebus FG, Sluiter JK, Kuijer PPF, Willems JH, Frings-Dresen MH. Work-ability evaluation: A piece of cake or a hard nut to crack? Disabil Rehabil. 2007;29(16):1295-1300.
6. Anner J, Schwegler U, Kunz R, Trezzini B, de Boer W. Evaluation of work disability and the international classification of functioning, disability and health: what to expect and what not. BMC Public Health. 2012;12(1):470.
7. Louwerse I, Huysmans MA, van Rijssen HJ, van der Beek AJ, Anema JR. Characteristics of individuals receiving disability benefits in the Netherlands and predictors of leaving the disability benefit scheme: a retrospective cohort study with five-year follow-up. BMC Public Health. 2018;18(1):157.
8. Kok R, Hoving JL, Verbeek J, Schaafsma FG, van Dijk FJ. Integrating evidence in disability evaluation by social insurance physicians. Scand J Work Environ Health. 2011;37(6):494-501.
9. Ilmarinen J. Work ability—a comprehensive concept for occupational health research and prevention. Scand J Work Environ Health. 2009;35(1):1-5.
10. Alavinia SM, De Boer A, Van Duivenbooden J, Frings-Dresen M, Burdorf A. Determinants of work ability and its predictive value for disability. Occup Med. 2009;59(1):32-37.
11. De Boer A, Verbeek J, Spelten E, Uitterhoeve A, Ansink A, De Reijke T, et al. Work ability and return-to-work in cancer patients. Br J Cancer. 2008; 98(8):1342-1347.
12. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ. 2005;330(7494):765.
13. Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med. 2012;157(1):29-43.
14. King G, Zeng L. Logistic Regression in Rare Events Data. Political Analysis. 2001;9(2):137-163.
15. Wei WWS. Time Series Analysis, in The Oxford Handbook of Quantitative Methods in Psychology, T.D. Little, Editor. 2013, Oxford University Press: New York. p. 458-485.
16. Levy PS, Lemeshow S. Sampling of Populations: Methods and Applications. 4th ed. 2008, New Jersey: John Wiley & Sons.
17. Weerdesteijn KH, Schaafsma FG, Louwerse I, Huysmans MA, Van der Beek AJ, Anema JR. Does self-perceived health correlate with physician-assessed functional limitations in medical work disability assessments? J Psychosom Res. 2019;125:109792.
18. Ilmarinen J, Tuomi K. Work ability of aging workers. Scand J Work Environ Health. 1992;18(2):8-10.
19. Ahlstrom L, Grimby-Ekman A, Hagberg M, Dellve L. The work ability index and single-item question: associations with sick leave, symptoms, and health – a prospective study of women on long-term sick leave. Scand J Work Environ Health. 2010;36(5):404-412.

20. El Fassi M, Bocquet V, Majery N, Lair ML, Couffignal S, Mairiaux P. Work ability assessment in a worker population: comparison and determinants of Work Ability Index and Work Ability score. BMC Public Health. 2013;13(1):305.
21. Boström M, Sluiter JK, Hagberg M. Changes in work situation and work ability in young female and male workers. A prospective cohort study. BMC Public Health. 2012;12(1):694.
22. Ekman A, Ahlstrand C, Andrén M, Boström M, Dellve L, Eriksson J, et al. Ung Vuxen—Basenkät. [Young Adults—Baseline Questionnaire]. Report from the Department of Occupational and Environmental Medicine, Nr 118. 2008, Gothenburg University: Gothenburg.
23. SSI. Het Claimbeoordelings- en Borgsysteem; Een introductie voor belangstellenden. [The Claim Assesment and Monitoring System; an introduction for interested people]. 2003, Dutch Social Security Institute: Amsterdam.
24. Ware Jr JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. Med care. 1992;30(6):473-483.
25. Pilowsky I. Dimensions of Hypochondriasis. Br J Psychiatry. 2018;113(494):89-93.
26. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand. 1983;67(6):361-370.
27. Vendrig AA, Schaafsma FG. Reliability and Validity of the Work and Well-Being Inventory (WBI) for Employees. J Occup Rehabil. 2018;28(2):377-390.
28. Kaymaz I, McMahon CA. A response surface method based on weighted regression for structural reliability analysis. Probabilist Eng Mech. 2005;20(1):11-17.
29. Kennedy P. A Guide to Econometrics. 6th ed. 2008, West Sussex: John Wiley & Sons.
30. Akobeng AK. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. Acta Paediatr. 2007;96(3):338-341.
31. Siegel SC, Castellan J. Nonparametric statistics for the behavioural sciences. 1988, New York: McGraw-Hill.
32. Feldt T, Hyvönen K, Mäkikangas A, Kinnunen U, Kokko K. Development trajectories of Finnish managers' work ability over a 10-year follow-up period. Scand J Work Environ Health. 2009;35(1):37-47.
33. Van Muijen P, Duijts SFA, Bonefaas-Groenewoud K, van der Beek AJ, Anema JR. Predictors of fatigue and work ability in cancer survivors. Occup Med. 2017; 67(9):703-711.
34. Van den Berg T, Elders L, de Zwart B, Burdorf A. The effects of work-related and individual factors on the Work Ability Index: a systematic review. Occup Environ Med. 2009;66(4):211.
35. Radkiewicz P, Widerszal-Bazyl M. Psychometric properties of Work Ability Index in the light of comparative survey study. Int Congr Ser. 2005;1280:304-309.
36. Burdorf A. Prevention strategies for sickness absence: sick individuals or sick populations? Scand J Work Environ Health. 2019;45(2):101-102.
37. Pedersen J, Gerds TA, Bjorner JB, Christensen KB. Prediction of future labour market outcome in a cohort of long-term sick-listed Danes. BMC public health. 2014;14(1):494.
38. Roelen C, Thorsen S, Heymans M, Twisk J, Bültmann U, Bjørner J. Development and validation of a prediction model for long-term sickness absence based on occupational health survey variables. Disabil Rehabil. 2018;40(2):168-175.
39. Dionne CE, Bourbonnais R, Frémont P, Rossignol M, Stock SR, Larocque I. A clinical return-to-work rule for patients with back pain. Can Med Assoc J. 2005; 172(12):1559-1567.
40. Heymans MW, Anema JR, van Buuren S, Knol DL, van Mechelen W, De Vet HC. Return to work in a cohort of low back pain patients: development and validation of a clinical prediction rule. J Occup Rehabil. 2009;19(2):155-165.

41. Nieuwenhuijsen K, Verbeek JH, de Boer AG, Blonk RW, van Dijk FJ. Predicting the duration of sickness absence for patients with common mental disorders in occupational health care. Scand J Work Environ Health. 2006;32(1):67-74.
42. Heijbel B, Josephson M, Jensen I, Stark S, Vingård E. Return to work expectation predicts work in chronic musculoskeletal and behavioral health disorders: prospective study with clinical implications. J Occup Rehabil. 2006; 16(2):169-180.
43. Sampere M, Gimeno D, Serra C, Plana M, López JC, Martínez JM, et al. Return to work expectations of workers on long-term non-work-related sick leave. J Occup Rehabil. 2012;22(1):15-26.
44. Cohen EL, Caburnay CA, Luke DA, Rodgers S, Cameron GT, Kreuter MW. Cancer coverage in general-audience and black newspapers. Health Commun. 2008;23(5):427-435.
45. Koh HC, Tan G. Data mining applications in healthcare. J Healthcare Inf Manage. 2011; 19(2):65.

CHAPTER 5

Preferences regarding the way of use and design of a work ability prognosis support tool: a focus group study among professionals

I Louwerse
MA Huysmans
HJ van Rijssen
J Overvliet
AJ van der Beek
JR Anema

Published in:
Disability and Rehabilitation
2019 Nov 26:1-7

Abstract

Purpose To explore the preferable way of use and design of a work ability prognosis support tool for insurance physicians (IPs) and labour experts (LEs), based on a prediction model for future changes in work ability among individuals applying for a work disability benefit.

Methods We conducted three focus groups with professionals of the Dutch Social Security Institute (17 IPs and 7 LEs). Data were audio recorded and qualitatively analysed according to the main principles of thematic analysis.

Results Clarity and ease of use were mentioned as important features of the tool. Most professionals preferred to make their own judgement during the work disability assessment interview with the claimant and afterwards verify their evaluation with the tool. Concerning preferences on the design of the tool, dividing work disability claimants into categories based on the outcome of the prediction model was experienced as the most straightforward and clear way of presenting the results. Professionals expected that this encourages them to use the tool and act accordingly.

Conclusions The tool should be easy to access and interpret, to increase the chance that professionals will use it. This way it can optimally help professionals making accurate prognoses of future changes in work ability.

Implications for rehabilitation

- A work ability prognosis support tool based on a prediction model for changes in work ability at one-year follow-up can help occupational health professionals in making accurate prognosis of individuals applying for a work disability benefit.
- To be used in occupational health practice, these tools should have a simple and easy-to-use design.
- Graphical risk presentation can be used to provide intuitive meaning to numerical information and support users' understanding.
- Taking professionals' preferences into account when developing these tools encourages professionals to use the tools and act accordingly.

Background

Individuals who are unable to work due to a disease or disorder can apply for a work disability benefit. In most European countries, this covers both financial support to compensate loss of income, and interventions to support return to work [1]. Insurance physicians (IPs) and labour experts (LEs) in the Netherlands assess disorders and functional abilities to determine whether a work disability benefit should be granted.

In many countries, prognosis of future changes in work ability is an important task of medical doctors during the medical disability assessment because, once a work disability benefit has been granted, changes in health may alter continuing eligibility [1, 2]. Medical re-assessments are conducted to determine whether an individual's health has improved or deteriorated to such an extent that adjustment of the benefit or support to return to work is required. In the Netherlands, IPs need to determine during a medical disability assessment if and when a re-assessment should be planned. To ensure that these medical re-assessments are planned at the time an assessment interview with an IP or LE has most added value, accurate prognosis of work ability is important. In general, claim duration for work disability benefits is long lasting for many claimants [3, 4]. Because long-term occupational inactivity is bad for an individual's health, and returning to work is generally associated with a positive effect on the future course of the disease and work ability, accurate prognosis of future changes in work disability may also enable effective return to work support [5, 6].

However, making a prognosis of the future course of work ability is not that easy, because it requires rather complex predictions, in which a broad range of factors play a role. These include common sociodemographic and health-related characteristics such as age, educational level and diagnoses, but also more subjective measures such as coping strategies, health experience and social support from relatives [7]. We developed a statistical model predicting future changes in work ability based on a broad range of factors, selected by a literature search of potential prognostic variables [8]. The outcome of the prediction model is for each individual claimant the likelihood of the change in work ability one year later.

The aim of the prediction model is to help professionals in making an accurate prognosis of work ability for individual claimants during the work disability assessment. In order to make the results of the prediction model easily accessible and interpretable for professionals, it needs to be supported by a suitable interface, in medicine often called clinical decision support tools. These are developed to support decision-making, in which the characteristics of individual patients are matched with a clinical knowledge base or decision rule [9]. The tools present patient-specific assessments or recommendations to clinicians at the time they have to make the decisions. Clinical decision-support tools are designed to aid decision-making; they can introduce efficiencies into the system, optimize the time with the client, and improve the overall quality of services and return to work interventions [10, 11]. In medical practice, clinical decision support tools can increase health care quality and efficiency [10, 12]. To be effective, not only the evidence base underlying these tools needs to be relevant and of high quality, but also the tool itself should be easily accessible and interpretable. Anticipating professionals' needs on the preferred way of use and design are key components when developing effective and implementable decision support tools [12].

Prognosis of work ability is an important task of IPs and LEs, and an evidence-based prediction model for future changes of work ability can help them making accurate prognosis. Although such tools are more common in clinical practice, they are currently lacking in work disability assessments. In order to develop a useful and relevant work ability prognosis support tool via which the outcome of the prediction model can be provided to IPs and LEs, it is important to know how and where in the decision-making process these professionals will use the tool and how they like the results to be presented. Therefore, the aim of this study was to obtain information on the preferences of IPs and LEs regarding the way of use and design of a work ability prognosis support tool regarding the prognosis of work ability of disability benefit claimants', based on a prediction model.

Methods

This study employs a qualitative focus group approach to explore the preferable way of use and design of a work ability prognosis support tool based on a prediction model for future changes in work ability. Unlike individual interviews, focus groups allow for interaction among group members. This enhances creativity, and makes it a useful approach for generating ideas, attitudes and opinions about a topic. Conducting focus groups helps individual participants to become aware of the range of design and use options and possible ways in which the work ability prognosis support tool can be applied. We conducted three focus groups with IPs and LEs working at the Dutch Social Security Institute (SSI). IPs are medical doctors who conduct disability assessments, based on diagnoses and functional abilities. Subsequently, LEs, who often have a background in social work, conduct an assessment of corresponding job opportunities. According to Dutch law (WMO), no ethical approval was necessary for this study, because no patients were included in the study and the physicians were not exposed to any intervention.

Work ability prognosis support tool

The goal of the focus groups was to examine the usability of a work ability prognosis support tool, based on an evidence-based prediction model that identifies claimants with a high probability of experiencing an improvement in Work Ability Score (WAS) at one-year follow-up [13]. The prediction model was based on a longitudinal cohort of 944 individuals, who were granted a work disability benefit by the SSI. Statistical variable selection was used to select the prognostic factors that were included in the final model. These were several physical and mental functioning factors, work status, wage loss, and work ability at baseline. The outcome of the prediction model is, for each individual claimant, the expected change in work ability at one-year follow-up. This outcome can be used by IPs and LEs as an additional source of information when they need to make decisions about the prognosis of claimants applying for a work disability benefit. For more information on the prediction model, we refer to Louwerse et al [8]. The participants of the focus groups did not have any knowledge in advance about the prediction model, but a short presentation about the development and prognostic factors was given at the start of the focus group meetings.

Participants

The SSI has 27 offices, divided over 12 regions. In total, about 900 IPs and LEs were working at the SSI in 2018. Each focus group consisted of both IPs and LEs, who were working in the same region but possibly at different offices. This was done to reduce travel time, thereby making it easier to participate, and because there are very small differences in work procedures between regions. Convenience sampling was used to recruit participants. To start with, three regions were selected based on their willingness to participate and geographical distribution; one in the west, one in the middle, and one in the east of the Netherlands. Depending on the level of data saturation, i.e. whether new themes did emerge when analysing the third focus group, more focus groups could be organized. IPs and LEs were recruited via their supervisors. A prerequisite was that all participants currently need to perform medical disability assessment interviews. In order to capture a wide range of perspectives on the preferable way of use and design of the work ability prognosis support tool, we informed the supervisors that we aimed at a range of demographic characteristics (gender, age) and years of working experience when recruiting the participants. However, recruitment of participants was voluntary and all IPs and LEs who showed interest were accepted. Participants received no compensation for their participation in the focus groups.

Both IPs and LEs are involved in prognosis of future changes of work ability. However, while for IPs it is one of their main tasks during the medical disability assessment, LEs in the Netherlands mainly focus on current limitations and corresponding job opportunities. As IPs will be the main users of the work ability prognosis support tool, we aimed for at least two thirds of the participants being IPs.

Data collection

The focus groups were held in May and June 2018, at an office in the region where the IPs and LEs were based. One moderator (IL) and one observer (MHA and HJvR), all working as researchers in the field of occupational health, facilitated the focus groups. JO was present at all focus groups to take notes. For IL, it was the first time as a moderator. However, all other researchers (MHA, HJvR, JO, AvdB and JRA) had previous experience with conducting focus group meetings. Moreover, the procedures and topics of the focus groups were discussed in detail in the research team beforehand. There were no established relationships between the moderator and the participants prior to the study.

All three focus groups lasted for about 1.5 hour, with a short break halfway through the focus group. The focus groups started with an explanation of the goal of the study, and the role of the moderator and the observer(s). Then, all participants introduced themselves and the topics were discussed. Two topics were discussed: 1) the preferred way of use of the work ability prognosis support tool, and 2) the preferable design of the work ability prognosis support tool. A more detailed overview of the topics is provided in table 1. The topic guide for the focus groups was developed based on extensive discussions during several meetings of the research team.

Two weeks before the focus group meeting, participants were sent an information letter stating the goal and procedures of the focus groups, and the data management process. Participants were informed that everything discussed during the focus group would be handled confidentially, and all quotes would be anonymized. If participants agreed, they were asked to sign the informed consent form that was enclosed. During the meeting, data were recorded with an audio-recording device. Besides, the observer took notes of the topics discussed. Before the start of the focus group, participants were asked to fill in a short questionnaire, regarding demographics and working experience. Each participant was then given a number, linked to the questionnaires, and their names were not used in the analysis. Within one week after the meeting, participants received a summary of the content of the focus group, which they were asked to check. They were asked to contact the researchers if they found any errors or omissions. In the results, we used quotes originating from the interviews to illustrate our findings. Cited professionals were described by the job title of their profession, gender and age. Quotes were translated by one researcher (IL) and checked by all other researchers.

Analysis

Data were analysed using a thematic analysis approach [14]. The COREQ checklist for reporting qualitative studies was used [15]. All data were transcribed verbatim in Dutch. The data collection continued until saturation of information was established, i.e. the transcripts of the meetings provided no new information. The focus groups were analysed according to the main principles of thematic analysis, i.e. through a systematic classification process of coding and identifying themes or patterns in order to describe the preferable way of use and design of the work ability prognosis support tool [16, 17]. All transcribed text and the notes of the focus groups were used in the analysis. First, two researchers (IL and JO) coded two thirds of each of the focus groups transcripts independently. During this phase of open coding, transcripts were carefully read, text parts that seemed relevant were coded and relations between main and sub codes were suggested. Second, during the phase of axial coding, the researchers discussed whether the created codes were appropriate to describe the data and whether the relation between main and sub codes was appropriate. This discussion continued until consensus was reached. As the researchers were both present during all focus groups, the interaction between the participants was taken into account as well. Finally, patterns in the data were identified by looking for returning themes and by making connections between these themes. After consensus was reached, all transcripts were (again) analysed by IL, using the provisional code list. All analyses were conducted using ATLAS.ti software.

Results

Participants

After three focus groups, a satisfactory level of data saturation was reached and therefore no additional focus groups were organized. In total 24 professionals participated; 5 in the first focus group (3 IPs and 2 LEs), 8 in the second (6 IPs and 2 LEs), and 11 in the third (8 IPs and 3 LEs). The actual distribution of 17 IPs and 7 LEs in total was in line with the intended distribution of about two thirds of the participants being IP, and one third being LE. The mean age of the participants was 51 years (SD = 9 years), 9 were female and 15 were male. The average working experience was 17 years (SD = 12 years) for IPs and 14 years (SD = 6 years) for LEs.

Table 1. Overview of focus group topics

Topics for IPs and LEs
1. Preferred way of use of the work ability prognosis support tool
a. Moment of use during the medical disability assessment
i. Before the interview with the claimant (when preparing for the interview)
ii. During the interview with the claimant (real-time interaction with the claimant)
iii. After the interview with the claimant (to verify or falsify own evaluation)
b. Reasons for non-use: situations in which the prediction model and the work ability prognosis support tool might not be informative
2. Preferable design of the work ability prognosis support tool
a. Way of presenting the outcome of the prediction model, e.g. on a continuous scale or dividing claimants into categories
b. Presenting additional information on the uncertainty of the outcome of the prediction model, e.g. by showing confidence interval or a larger number of categories

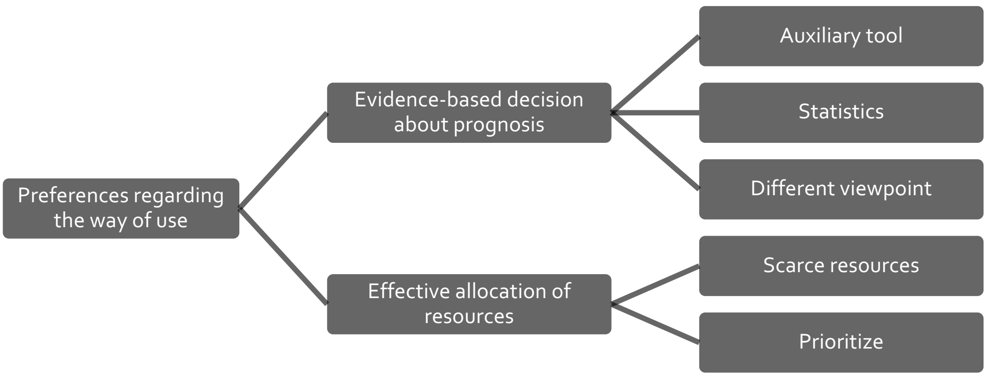


Figure 1. Part of the coding tree of the thematic analysis

Only one participant added a remark to the content of the summary provided for them to check. The participant stressed the importance of a certain issue mentioned during the focus group. The comments of this participant were taken into account in the results.

Our findings are presented per theme: the preferences regarding the way of use are set out first, followed by the preferences regarding the design of the decision support tool. Quotes were used to illustrate our findings, whereby we aimed for a distribution of profession, gender and age that represents the distribution of these characteristics among the participants. As an example of the coding tree that we developed, figure 1 shows part of the coding tree for the first theme.

Preferences regarding the way of use

Evidence-based decisions about prognosis

Accurate prognosis of future work ability was considered as an important task of IPs and LEs. It aids provision of effective interventions to return to work for claimants who will benefit most from it. Assisting claimants to get back to work is stated in the vision of the SSI, and participants mentioned it as an important motivator for work.

Some participants had difficulties determining how to value the prediction model compared to their own consideration and estimation of a claimant's prognosis. In answer to this, other participants argued that it aims to be an auxiliary tool, based on statistics of large numbers rather than physician's knowledge and experience, which gives a different viewpoint and helps professionals to make more evidence-based prognosis.

> "*It is helpful as an auxiliary tool: is the prognosis that I make based on literature and guidelines correct?*" (Insurance physician, female, 40 years)

Effective allocation of resources

Concerning reasons to use the work ability prognosis support tool, participants thought that, in case of scarce resources, the work ability prognosis support tool could guide effective allocation of resources. IPs conduct medical re-assessments to determine whether an individual's health and work ability have improved or deteriorated to such an extent that adjustment of support to return to work is required. The participants suggested that the work ability prognosis support tool can help to select the claimants who are most likely to experience a significant improvement in work ability and hence are expected to benefit most from return to work support. According to the participants, the tool assists professionals to prioritize these claimants when they need to allocate limited resources and plan re-assessments.

> "*It is also possible to prioritize, depending on the available capacity. We first assign resources to claimants who are most likely to benefit from it. Depending on the capacity that remains, we can assign it to claimants of whom we are less sure*." (Labour expert, male, 56 years)

Verify and validate own prognosis

Opinions on when to use the tool differed. The majority of the participants stressed that they are open-minded at the start of the work disability assessment interview. They felt that using the work ability prognosis support tool at this stage would prohibit them from being so. They would first want to make their own judgement during the interview with the claimant and afterwards verify or falsify their evaluation with the work ability prognosis support tool. It will strengthen their belief that they made the right decision if the outcome of the prediction model matches their own prognoses, and it will be a reason to reconsider their prognoses if it does not match. A prerequisite in this case is that the work ability prognosis support tool should be available briefly after the work disability assessment interview, when IPs need to write a medical report.

> "*Then I can first get my own idea and afterwards see if this coincides with the outcome of the prediction model. If you consult the work ability prognosis support tool before the interview with the claimant, then it might be in the back of my mind during the interview. Although it will not completely determine my way of thinking, it might influence it anyway, and cause you to ask certain questions that you would otherwise not have asked*." (Insurance physician, male, 58 years)

More thorough preparation for the interview

A few participants argued that being able to use the work ability prognosis support tool before the disability assessment interview would help in preparing the interview, and might give guidance for topics to pay additional attention to during the interview with the claimant. Moreover, it gives the possibility to discuss the prognosis and follow-up actions with the claimant during the interview.

> "*I think it could be useful when preparing for the interview with the claimant. What information is, for instance, available about expected return to work and the motivation of the claimant? I could use this information to ask specific questions to the claimant and talk about potential barriers for return to work.*" (Insurance physician, male, 41 years)

No distraction from the conversation with the claimant

All participants agreed that using the work ability prognosis support tool during the disability assessment interview is not desirable, as during this contact they want to focus full attention on the interview with the claimant. Using the work ability prognosis support tool and judging the outcome of the prediction model would distract them and would take too much time.

> "*You need time to interpret and evaluate the work ability prognosis support tool. It is difficult to do this during the interview with the claimant*." (Insurance physician, male, 60 years)

Reservations about self-reported factors

Participants expressed their concern about using self-reported measures as prognostic variables for the prediction model. They mentioned that applying for a work disability benefit is an emotional process during which claimants can experience many insecurities and that these emotions could influence the answers claimants give on self-reported questionnaires.

Moreover, they questioned whether the work ability prognosis support tool would be for professional use only, or if and how its results can be communicated to claimants as well.

> "*We have claimants who have negative thoughts about getting back to work, while we think that there are opportunities for return to work. If claimants own opinions greatly influence the outcome of the prediction model, it would be less reliable for me*." (Labour expert, male, 53 years)

One model for all claimants

Concerning reasons for non-use, participants mentioned some factors that by themselves almost completely determine the expected prognoses and the recommended return to work interventions were mentioned. For instance, older claimants often have less psychological resilience than younger claimants, and less motivation to put a lot of effort in a return to work intervention when they are close to reaching the retirement age. Other factors that were mentioned as determinants for lower expectations of future improvements of work ability and successful return to work support were claimants suffering from comorbidity, claimants with non-health related complications, such as financial or personal issues, and claimants with several and longer periods of work disability in the past.

> "*Age plays a role. For instance when a woman older than 50 years has a long-term depression, her psychological resilience becomes less and you can take that into account during your assessment*." (Insurance physician, male, 58 years)

> "*Claimants who are already longer without a job, either because of unemployment or sickness, have a larger distance to the labour market, and are usually also less motivated.*" (Labour expert, female, 58 years)

However, participants agreed that they would not beforehand exclude participants from the prediction model. The work ability prognosis support tool has most added value in cases where professionals are unsure about the prognosis, for instance when the course of the disease is unclear. However, also when an IP is more confident about the expected change in work ability, some of them argued that they could use the work ability prognosis support tool in these cases to verify their thoughts. Because the prediction model is based on a diverse set of variables, both SSI registration data and self-reported measures, it is a useful complement to the prognosis of the professional.

> "*I think it always adds something, because when I think an improvement of work ability is very likely and the work ability prognosis support tool shows green, then this confirms my expectations. Otherwise, the tool shows the opposite, it would make me think I missed*

something, I will think about it more thoroughly. So I think it is always useful." (Insurance physician, male, 59 years)

Preferences regarding the design

Clearly present the outcome of the prediction model

Participants mentioned clarity and ease of use as important features of the tool. They argued that these features are crucial for successful use of the tool in practice. Dividing claimants into groups based on their predicted future change in work ability and assigning colours to the groups (e.g. green for claimants with a high probability of experiencing an improvement in work ability, red for claimants with a high probability of experiencing a deterioration in work ability, and orange if no relevant change in work ability is predicted) was seen as a straightforward and clear way of presenting the results of the prediction model. As users have a quick and uniform association with colours, this encourages them to take action.

"*By using colours in the outcome of the prediction model it is immediately clear.*" (Labour expert, female, 47 years)

"*I also think that if you represent it this way, that it encourages us to take action faster than if you would represent it in a more neutral way.*" (Insurance physician, male, 31 years)

Detailed information about the predicted outcome

Participants argued that more detailed information might be helpful in some cases, for instance when for a claimant the predicted value falls on the boundary of two categories. More precise presentation of the outcome of the prediction model, e.g. by dividing the claimants into more than three categories or by presenting the prediction on a continuous scale, gives more accurate information but would also be more difficult to interpret. Moreover, it would result in less uniformity as information that is more detailed leads to larger differences in interpretation between professionals. Moreover, some participants mentioned that they would like to have some information about the uncertainty of the predicted value (e.g. the outcome of the prediction model and the corresponding category that a claimant is assigned to), for instance by presenting the confidence interval around the predicted value on a continuous scale. However, most participants thought that it would be difficult for them to value this information and that it would work against an easy interpretation of the main outcome.

"*A continuous scale is visually attractive, but if you want to evaluate it, you should attach recommendations or actions to it. Otherwise everyone will interpret it in his or her own way, and that will not result in better prognosis*." (Insurance physician, female, 40 years)

Information about the prognostic variables

To create more support and confidence in the work ability prognosis support tool and the underlying prediction model, participants argued that they need more information about the prognostic variables of the prediction model. Some participants preferred a concise summary of characteristics of claimants and the prognostic variables of the prediction model, for other participants a presentation beforehand about how the prediction model is constructed and its underlying variables would be sufficient.

> "*It is difficult for me to interpret the outcome of the prediction model if I don't have information about what the model looks like. What are the most important factors? Some general information about the prediction model and the variables that are included would be helpful.*" (Labour expert, female, 47 years)

> "*Also for the acceptance of my colleagues it is important that we have a better understanding of the prediction model.*" (Insurance physician, male, 47 years)

Discussion

Main findings

The goal of this study was to explore the preferences of professionals regarding the way of use and design of a work ability prognosis support tool, which can help them in making accurate prognoses of future changes in work ability. Qualitative analysis of focus groups showed that IPs and LEs of the SSI reported a large number of preferences regarding the way of use (e.g. evidence-based decisions about prognosis, effective allocation of resources, and verify and validate own prognosis) and preferences regarding the design (e.g. clearly present the outcome of the prediction model and information about the prognostic factors) of the decision support tool. Participants agreed that clarity and ease of use are important features of the tool. Dividing claimants into categories based on the outcome of the prediction model and assigning colour labels to the classes was experienced as the most straightforward and clear way of presenting the results of the prediction model. It encourages professionals to use the tool and act accordingly. Concerning preferences on when to use the tool, most professionals would prefer to first make their own judgement during the work disability assessment interview with the claimant and afterwards verify or adjust their evaluation based on the outcome of the work ability prognosis support tool.

Interpretation of findings

Ease of use was mentioned by the participants of the focus groups as key component for successful actual use of the tool. This is in agreement with previous qualitative research that concluded that a simple and easy-to-use design was a necessity for IPs to use a prediction rule aimed at supporting work disability assessment of cancer survivors [18]. In this earlier study, IPs mentioned that a prediction rule should take little time to use, should have added value for the work disability assessment and should be both valid and reliable. Ensuring ease of use by using computers to generate support is an important feature for clinicians and other health care stakeholders to use clinical support tools [19]. Even if the tool is very well designed, it will be useless if the professional is not able to use prediction rule at the time of decision-making [20].

Another crucial component for successful use of the tool mentioned by the participants of our focus groups was clarity of the work ability prognosis support tool. This concurs with the results of studies describing the use of computer tools among physicians, stating that professionals will not be happy about using a support tool if the information that it presents does not fit on a single screen [21]. Our participants indicated that even simple and relatively straightforward recommendations could be interpreted in different ways, depending on one's perspective or experience. This is in line with results found in the general field of risk communication that people have difficulties with interpreting and providing meaning to numerical information [22, 23]. Even across samples consisting of only highly educated individuals, participants appeared to have difficulties understanding and interpreting health statistics [24, 25]. Instead, graphical risk presentation, such as using colours and verbal categorical labels, intends to provide intuitive meaning to numerical information. By using graphical formats, it is usually easier to attract the attention of the user and to support their understanding [26, 27]. Dividing claimants into three categories based on the prediction model and assigning colours and recommendations to these categories seems a good way to prevent differences in interpretation.

The present study revealed that most SSI professionals would like to use the work ability prognosis support tool to verify their own prognosis directly after the work disability assessment interview. A minority of the participants mentioned that they would prefer to have the possibility to use the tool before the disability assessment interview as well, as this would help them in preparing the interview with the claimant. In general, automatic provision of decision support at the time and location of decision-making is a key element for successful actual use [28]. It would be possible to give SSI professionals the opportunity to use the tool at other moments as well, but it should be investigated whether professionals would do so, as previous research showed that clinical support tools were less successful if clinicians had to initiate the use of it themselves [19].

Strengths and limitations

A first strength of the present study is that the focus groups were held within three different regions of the SSI, each located in a different part of the Netherlands, and the participants were selected out of all potential users of the work ability prognosis support tool. This enabled us to gather different perspectives, design a tool that satisfies as much as possible the demands, and create a high level of support for the users. Secondly, we carefully followed the guidelines for qualitative research by having two researchers who independently analysed and coded the data, by discussing the study design and results within the research team, and by performing member checks. This contributes to the credibility and confirmability of the study. The dependability of the study was taken into account by continuing data collection until a satisfactory level of data saturation was reached, and by applying a flexible research design.

Another strength of this study is that, when considering the way of use and design of the work ability prognosis support tool, we focused on its actual use in the Dutch social security system and the procedures at the SSI. Hence, the setting in which the work ability prognosis support tool will be used is emphasized during the focus groups. However, the obvious limitation related to this is that generalizability of our findings was limited. Translation of our findings to other settings can only be done with care as work disability evaluation processes and legislation largely differ across European countries in terms of steps involved, use of professional assessors and time consumption [1]. However, by giving a description of social security context in which this study was performed, and by comparing our findings with existing literature we aimed at enhancing the transferability. Other limitations of this study are that we conducted convenience sampling, which means that the participants might have provided only limited different perspectives because they were not sampled by purposeful sampling. A final limitation is that we did only focus on the preferable way of use and design of the work ability prognosis support tool, and that there was no time to discuss other barriers or facilitators for use that could be faced in practice. Therefore, it is important to conduct a process evaluation alongside an effectiveness study to identify other barriers or facilitators for use.

Implications for practice and research

A work ability prognosis support tool based on a prediction model for changes in work ability at one-year follow-up can help IPs and LEs making accurate prognosis. Being supported by a suitable work ability prognosis support tool, which is easy to access and interpret, is a prerequisite to increase the chance that professionals will use the tool. The present study showed that IPs and LEs agree on the preferred way of use and design of the work ability prognosis support tool. This provides a good starting point for developing a tool that is user-friendly and aligned to the preferences of IPs and LEs and that can be tested in a trial. Based on the results of this focus group study, we will develop a work ability prognosis support tool. Next, an effectiveness study will be performed to determine if the actual use of the tool contributes to more accurate prognoses. Furthermore, a process evaluation should show whether IPs actually use the tool and how they evaluate it.

Although the development of clinical support tools has rapidly increased over the past decade, it remains to be seen whether these tools will be part of everyday practice and to what extent they can contribute to effective occupational health provision. As clinical support tools do not guarantee a correct solution for every single case, it is important to emphasize that they should not be automatically followed. Rather, these tools should be complementary to occupational health professionals' judgements, which should be prioritized at all times. Therefore, professionals should be informed about proper use and the scientific evidence of such tools.

References

1. De Boer W, Besseling JJM, Willems JHBM. Organisation of disability evaluation in 15 countries. Prat Organ Soins. 2007;3(38):205-217.
2. Anner J, Kunz R, de Boer W. Reporting about disability evaluation in European countries. Disabil Rehabil. 2014;36(10):848-854.
3. Louwerse I, Huysmans MA, van Rijssen HJ, van der Beek AJ, Anema JR. Characteristics of individuals receiving disability benefits in the Netherlands and predictors of leaving the disability benefit scheme: a retrospective cohort study with five-year follow-up. BMC Public Health. 2018;18(1):157.
4. Wise DA. Social security programs and retirement around the world: Historical trends in mortality and health, employment, and disability insurance participation and reforms. 2012, Chicago: University of Chicago Press.
5. OECD, Sickness, Disability and Work: Breaking the Barriers: A Synthesis of Findings across OECD Countries. 2010, Paris: OECD Publishing.
6. Waddell G, Burton AK. Is work good for your health and well-being? 2006, London: The Stationery Office.
7. Kok R, Verbeek JA, Faber B, van Dijk FJ, Hoving JL. A search strategy to identify studies on the prognosis of work disability: a diagnostic test framework. BMJ open. 2015;5(5):e006315.
8. Louwerse I, Huysmans MA, van Rijssen HJ, Schaafsma FG, Weerdesteijn KHN, van der Beek AJ, Anema JR. Predicting future changes in work ability of individuals receiving a work disability benefit: weighted analysis of longitudinal data. Scand J Work Environ Health. 2020; 46(2):168-176.
9. Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, Tang PC. Clinical decision support systems for the practice of evidence-based medicine. J Am Med Inform Assoc. 2001;8(6):527-534.
10. Roshanov PS, Fernandes N, Wilczynski JM, Hemens BJ, You JJ, Handler SM, et al. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. BMJ. 2013;346:f657.
11. Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J Am Med Inform Assoc. 2003;10(6):523-530.
12. Bates DW, Cohen M, Leape LL, Overhage JM, Shabot MM, Sheridan T. Reducing the frequency of errors in medicine using information technology. J Am Med Inform Assoc. 2001;8(4):299-308.
13. Ahlstrom L, Grimby-Ekman A, Hagberg M, Dellve L. The work ability index and single-item question: associations with sick leave, symptoms, and health – a prospective study of women on long-term sick leave. Scand J Work Environ Health. 2010;36(5):404-412.
14. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77-101.
15. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care. 2007;19(6):349-357.
16. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res. 2005;15(9):1277-1288.
17. Boeije H. Analysis in qualitative research. 2009, London: Sage publications.
18. Van Muijen P, Duijts S, Kornet-Van der Aa D, Van der Beek A, Anema J. Work disability assessment of cancer survivors: insurance physicians' perspectives. Occup Med. 2015;65(7):558-563.
19. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ. 2005;330(7494):765.

20. Lee F, Teich JM, Spurr CD, Bates DW. Implementation of physician order entry: user satisfaction and self-reported usage patterns. J Am Med Inform Assoc. 1996;3(1):42-55.
21. Shojania KG, Yokoe D, Platt R, Fiskio J, Ma'Luf N, Bates DW. Reducing vancomycin use utilizing a computer guideline: results of a randomized controlled trial. J Am Med Inform Assoc. 1998;5(6):554-562.
22. Reyna VF, Brainerd CJ. The importance of mathematics in health and human judgment: Numeracy, risk communication, and medical decision making. Learn Individ Differ. 2007;17(2): 147-159.
23. Maviglia, SM, Zielstorff RD, Paterno M, Teich JM, Bates DW, Kuperman GJ. Automating complex guidelines for chronic disease: lessons learned. J Am Med Inform Assoc. 2003;10(2):154-165.
24. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Med Decis Making. 2001;21(1):37-44.
25. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. Psychol Sci Public Interest. 2007;8(2):53-96.
26. Waters EA, McQueen A., Cameron LD. Perceived risk and health risk communication. The Routledge Handbook of Language and Health Communication. 2014, Routledge London: New York. p. 47-60.
27. Trevena LJ, Zikmund-Fisher BJ, Edwards A, Gaissmaier W, Galesic M, Han PK, et al. Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. BMC Med Inform Decis Mak. 2013; 13(2):S7.
28. Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux P, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA. 2005;293(10):1223-1238.

CHAPTER 6

Use of a decision support tool on prognosis of work ability in work disability assessments: an experimental study among insurance physicians

I Louwerse
MA Huysmans
HJ van Rijssen
C Gielen
AJ van der Beek
JR Anema

Published in:
Journal of Occupational Rehabilitation
2020 Jun 11:1-12

Abstract

Purpose Assessment of prognosis of work disability is a challenging task for occupational health professionals. An evidence-based decision support tool, based on a prediction model, could aid professionals in the decision-making process. This study aimed to evaluate the efficacy of such a tool on Dutch insurance physicians' (IPs) prognosis of work ability and their prognostic confidence, and assess IPs' attitudes towards use of the tool.

Methods We conducted an experimental study including six case vignettes among 29 IPs. For each vignette, IPs first specified their own prognosis of future work ability and prognostic confidence. Next, IPs were informed about the outcome of the prediction model and asked whether this changed their initial prognosis and prognostic confidence. Finally, respondents reported their attitude towards use of the tool in real practice.

Results The concordance between IPs' prognosis and the outcome of the prediction model was low: IPs' prognosis was more positive in 72 (41%) and more negative in 20 (11%) cases. Using the decision support tool, IPs changed their prognosis in only 13% of the cases. IPs prognostic confidence decreased when prognosis was discordant, and remained unchanged when it was concordant. Concerning attitudes towards use, the wish to know more about the tool was considered as the main barrier.

Conclusion The efficacy of the tool on IPs' prognosis of work ability and their prognostic confidence was low. Although the perceived barriers were overall limited, only a minority of the IPs indicated that they would be willing to use the tool in practice.

Introduction

Individuals who are unable to work due to a disease or disorder can apply for a work disability benefit. In most European countries, this covers both financial support to compensate loss of income, and interventions to support return to work [1]. Occupational health professionals conduct work disability assessments to evaluate whether a benefit should be granted. One of the main tasks during this assessment is estimating prognosis of work ability [2]. Accurate prognosis is important to determine when an individual's work ability will improve or deteriorate to such an extent that adjustment of the benefit or support to return to work is required [3]. However, it is also considered the most difficult part of the work disability assessment, because it requires rather complex predictions, in which a broad range of individual characteristics and external factors play a role [4, 5].

A potential solution to this problem is to provide occupational health professionals with evidence-based decision support tools. These tools are comprised of software and designed to aid decision-making [6]. They match characteristics of individual claimants with a computerized knowledge base to generate patient-specific assessments or recommendations [7]. Although previous research has shown that such tools could be a means to achieve more accurate estimates of prognosis, they are usually based on a limited number of prognostic factors and are not 100% correct [8, 9]. In addition, other factors play a role in decision-making in the medicolegal setting of work disability assessments. Hence, decision support tools are not meant to take over the job of professionals, but to support them by providing objective estimates of outcome probabilities to complement their professional expertise, competencies and experience. For instance, a decision support tool based on an evidence-based prediction model for future changes in work ability could aid professionals during work disability assessments [10, 11]. The tool could help professionals to make more precise estimations of future work ability and could increase their prognostic confidence [12]. To establish the possible efficacy of the decision support tool on the complex decision-making processes during work disability assessments, insight in actual use of the tool and occupational health professionals' attitudes towards such a tool is needed.

In general, occupational health professionals recognize the potential usefulness of evidence-based decision support tools. However, previous studies have shown that adherence to the use of innovations in medical settings can be a difficult task to accomplish [13]. Many factors may influence the use of decision support tools in practice, such as lack of knowledge about the innovation, negative attitudes and beliefs towards the innovation, perceived lack of time, lack of motivation, and organizational constraints [14, 15]. Moreover, barriers operate on different levels: they can be related to the professional, the patient, the organization, or the social and cultural context [16]. Insight into the barriers and facilitators for use of the decision support tool is needed to be able to further develop the tool in line with professionals' needs and select an appropriate strategy for future implementation [17].

The objectives of our study were: (i) to evaluate the efficacy of the decision support tool on Dutch insurance physicians' (IPs) prognosis of work ability using case vignettes; (ii) to investigate whether use of the tool affects IPs' prognostic confidence; and (iii) to quantitatively assess the attitudes of professionals towards a decision support tool, and the perceived barriers and facilitators for use.

Methods

An experimental study including six case vignettes was conducted to answer the research questions.

Context

In the Netherlands, IPs conduct medical disability assessments to evaluate whether a work disability benefit should be granted. Once a work disability benefit has been granted, changes in work ability may alter its continuing eligibility. Therefore, prognosis of future changes in work ability is an important task of IPs [18]. IPs conduct medical re-assessments to determine whether a claimant's health has improved or deteriorated to such extent that adjusting the benefit and/or support to return to work is necessary. Re-assessments are not only an operational aspect of the disability system, but a means to monitor claimants' functional abilities. Depending on the situation of individual claimants, the term or the extent of the financial support of the benefit could be changed, or new rehabilitation interventions could be offered. Claimants could have interest in a certain outcome of a re-assessment. However, IPs are trained in objective assessment of functional limitations; IPs do not mainly focus their assessment on claimants' self-perceived health complaints and impairments but they use many other factors as well [19-21].

During the work disability assessment, IPs need to indicate for each claimant if and when a re-assessment should be planned. Because of the large number of work disability claimants and the limited capacity to perform re-assessments, accurate prognosis is important for efficient planning of medical re-assessments and adequate interventions to support return to work. An evidence-based prediction model could help IPs in making more accurate prognosis of individual claimants during the work disability assessment. The prediction model is a regression equation that uses some prognostic factors to predict for each claimant the expected change in work ability at one-year follow-up [10]. A cohort of 944 claimants who were granted a work disability benefit by the SSI was used to develop the prediction model and for internal validation of the model [21]. Work ability was measured using the Work Ability Score (WAS), a single item of the Work Ability Index (WAI) questionnaire that asks participants to compare their current work ability with their lifetime best on a 0–10 scale [22]. Higher scores indicate better work ability, and an improvement or a deterioration in WAS of at least two points is considered to be a relevant change likely to have an effect on return to work and work disability benefit [23, 24]. Based on the predicted change in WAS at one-year follow-up, claimants are divided into three groups: claimants with no relevant change, an improvement, or a deterioration in the expected level of work ability at one-year follow-up compared to baseline. The prognostic factors of the prediction model are several physical and mental functioning factors, work status, wage loss, and work ability at baseline.

In order to make the outcome of the prediction model easily accessible and interpretable for professionals it needs to be supported by a suitable interface, i.e. a decision support tool. Based on professionals' preferences regarding the way of use and design of the tool, we developed such a tool [11]. The tool uses claimant-specific information from self-reported questionnaires and registration data from the Dutch Social Security Institute (SSI) [21]. This information is matched with a decision rule, and presents the predicted change in work ability. IPs could use the tool as an auxiliary source of information to estimate prognosis of work ability, guide return to work interventions and plan re-assessments.

Study design

To assess the efficacy of using the decision support tool and IPs attitudes towards the tool, we conducted an experimental study including six case vignettes based on past work disability assessment reports [25, 26]. Advantages of vignette studies are that they are more realistic and less abstract than traditional survey questions, and that they can give insight into decision-making processes in an experimental setting [27, 28]. In consultation with three IPs working at the research department of the SSI and based on real patients' records, we constructed detailed descriptions of six claimants who were granted a work disability benefit. These descriptions included demographic factors, information about the last job (working hours, work demands), disorders, medical and non-medical treatments, and functional limitations. All variables that were included as prognostic factors in the prediction model were also presented to IPs in the case vignettes. Table 1 shows a summary of the six case vignettes. The case vignettes were presented to respondents for evaluation. Cases were selected in such a way that several factors believed to influence the judgement were varied, e.g. demographics, disease type, disease history, and predicted change in work ability at one-year-follow-up. As an example, we have added a more detailed description of one of the case vignettes to the Appendix. The six cases were considered to sufficiently represent the most important factors in work disability assessments and the possible outcomes of the prediction model. As we did not want to pose an unnecessary burden on the already limited resources at the SSI, we decided not to include any additional cases.

Study population

The study population consisted of IPs working at the SSI. The SSI is a semi-governmental organisation that assesses sickness absence and work disability benefit claims, takes care of benefit payments and provides reintegration support. Seven of the 27 offices of the SSI were selected based on their willingness to participate and geographical distribution: three in the North, two in the central part and two in the South of the Netherlands. At each office, a meeting was organised to inform IPs about the design and questionnaires of the experimental study, and to give them more information about the prediction model and decision support tool. Thereafter, IPs were asked if they were willing to participate in the study and were invited per e-mail to fill in an online questionnaire. Participation was voluntary. Inclusion criteria were being registered as an IP or following the postgraduate education in insurance medicine, and conducting medical disability assessment interviews of work disability claimants. All participants signed informed consent and all data were anonymized. The Medical Ethics Committee of VU University Medical Centre in Amsterdam, The Netherlands, has confirmed that ethics approval is not necessary, because the Medical Research Involving Human Subjects Act (WMO) does not apply to our study.

Questionnaire

The questionnaire consisted of three sections: a general section, evaluation of the six case vignettes, and a section to evaluate aspects for use of the tool. The general section included questions about demographics and professional characteristics such as age, gender, and work experience. Next, the six case vignettes were presented. For each vignette, IPs were asked about their prognosis of future changes in work ability based on a detailed description of demographic, work, health and psychological factors. Based on this prognosis, they were asked to specify duration of the period after which they wanted to plan a re-assessment and the expected change in work ability at one-year follow-up. In line with the outcome of the prediction model, the latter question had three answering categories: improvement, no change, or deterioration of a claimant's work ability. Moreover, IPs were asked to rate the level of prognostic confidence on a numerical rating scale (NRS), ranging from 0 (no confidence) to 10 (complete confidence). Next, the decision support tool was presented showing the prognostic factors and the outcome of the prediction model. The prognostic factors were also included in the detailed description of the case vignettes, but were presented again to enable IPs to judge the exact factors that were used to estimate the predicted outcome. IPs were asked to re-evaluate their prognosis using the same description of the case vignettes and the outcome of the prediction model. IPs indicated whether consulting the tool led them to change their prognosis or not, and if so in which direction (i.e. better or worse). An open-ended question gave respondents the opportunity to explain why they did or did not change their prognosis. Besides, to assess whether the tool made them more or less confident about their prognosis, respondents again indicated their prognostic confidence on a numerical rating scale.

To study future use of the decision support tool, IPs were asked if and how often they considered the tool to be of benefit during the medical disability assessment. Moreover, it was examined whether IPs would consider using a decision support tool in the future to support their prognosis during the work disability assessment, how they would want to use the tool, and in what situations or for which types of claimants they would not use it. Finally, 15 statements about barriers and facilitators for use of a decision support tool in real practice were presented. These statements originate from an existing validated questionnaire and assess constructs related to the society, the organization, the claimant, the future user of the decision support tool, and the tool itself [29]. A 5-point Likert scale ranging from “Fully disagree” to “Fully agree” was used to rate the extent of agreement with each statement.

Statistical analysis

Descriptive statistics were used to describe the demographic and professional characteristics of the IPs (means, standard deviations, percentages). We assessed the efficacy of the decision support tool on the IPs’ prediction of prognosis. Prognosis was defined as either concordant or discordant according to whether the prediction of the decision support tool was or was not equal to IPs’ own prediction of prognosis, i.e. before evaluating the decision support tool. To measure whether use of the decision support tool led IPs to change their prognosis, the number and percentages of IPs in the three answering options (i.e. prognosis of work ability remained unchanged, got worse, or got better after evaluating the decision support tool) were calculated. Differences between cases with different types of limitations (mental, physical, or both mental and physical) were assessed using chi-square tests. For each case vignette, we calculated the proportions of observed agreement between IPs’ prognosis and the outcome of the prediction model and used McNemar's test for paired proportions to compare agreement before and after evaluating the decision support tool. Multilevel analyses were used to assess changes in the level of prognostic confidence after consulting the decision support tool, taking into account that the data were clustered within IPs. Concerning the statements on barriers and facilitators for use, we calculated the percentage “agree” and “fully agree” for all statements to identify possible barriers and facilitators for use of the decision support tool. All analyses were performed in RStudio for Windows, version 0.99.902. The significance level of all statistical tests was set at $p<0.05$.

Table 1. Main characteristics of the six case vignettes

Vignette	Gender	Age	Job demands	Disorder(s)	Type of limitations	Functioning and treatment	WAS[a] at baseline
1	Male	45	Mostly physical	Paraplegia	Physical	ADL dependence	1
2	Female	29	Mental + physical	Chronic kidney disease	Physical	Waiting list for a kidney transplant	1
3	Male	34	Mostly physical	First episode psychosis	Mental	Hospitalized for treatment	4
4	Male	58	Mental + physical	Cerebral bleed	Mental + physical	Psychologist, physiotherapist	3
5	Female	38	Mostly mental	Depression, PTSD	Mental	Psychotherapy and EMDR	0
6	Female	34	Mostly mental	PTSD, multiple fractures	Mental + physical	Psychologist, physiotherapist	0

[a] WAS = self-assessed work ability score

Results

Twenty-nine IPs voluntarily participated in the study. This was just above the minimum number of 28 IPs that was determined in a sample size calculation as number needed to answer the research questions. Table 2 summarizes the demographic and professional characteristics of the respondents. The majority was female (59%) and worked as a registered IP (62%).

Prognosis without and with the decision support tool

In 82 (47%) cases, the prognosis of the IP without information from the decision support tool was concordant with the outcome of the prediction model (Table 3). The prognosis of the IP was more positive in 41% (n=72) and more negative in 11% (n=20) of the cases. Differences in prognosis occurred most often in cases with both mental and physical limitations (90%) and least often in cases with only physical limitations (26%).

For all cases where the prognosis of the IP was concordant with the outcome of the prediction model, IPs did not change their prognosis after evaluating information from the decision support tool. In 22% (n=20) of the cases where the prognosis was not concordant, the IP changed his or her prognosis after evaluating information from the tool. In the majority of the cases where the prognosis was changed, IPs considered the prognosis to be worse after evaluating the decision support tool (75%; n=15). Whether or not the IP changed his or her prognosis was independent of the type of limitations (p=0.34).

Table 2. Demographic and professional characteristics of the respondents (n=29)

	n	%	Mean	SD
Gender				
Male	12	41		
Female	17	59		
Age (years)			44	11
<35	7	24		
35-44	10	34		
45-54	4	14		
55+	8	28		
Type of IP				
Registered	18	62		
Postgraduate student	11	38		
Working experience as IP (years)			12	11
<5	10	34		
5-9	7	24		
10+	12	41		

Overall, the observed agreement between IPs' prognosis and the outcome of the prediction model increased from 47% to 58% after IPs evaluated the decision support tool. If we look at each of the case vignettes separately, we can see that for two of the vignettes that showed low initial agreement, there was an improvement in agreement after IPs were informed about the outcome of the prediction model. For these vignettes, the prediction model estimated no change in future work ability. There was a statistically significant change in the number of IPs that first predicted an improvement or deterioration in work ability and, after use of the tool, agreed with the outcome of the prediction model.

6

Confidence in the prognosis with and without the decision support tool

Table 4 presents the prognostic confidence of IPs without and with information from the decision support tool. In 57% (n=99) of the cases, the prognostic confidence of the IP changed after evaluation of the decision support tool. The confidence increased in 26% (n=45) and decreased in 31% (n=54) of the cases. Change in prognostic confidence occurred less often when the prognosis of the IP was concordant with the outcome of the decision support tool (n=40; 49%) than in case of discordant prognosis (n=59; 64%).

The results of the multilevel analyses showed that there was an overall decrease in prognostic confidence by 0.5 points on the NRS from 7.1 to 6.6 points (p=0.02). If the prognosis was concordant, the prognostic confidence increased from 7.2 to 7.4 points, but this increase was not statistically significant (p=0.26). The prognostic confidence significantly decreased from 7.0 to 6.0 points in cases with discordant prognosis (p<0.001).

Use of the decision support tool in practice

28% (n=8) of the respondents indicated that they would be willing to use an evidence-based decision support tool based on a prediction model during future work disability assessments. The majority of the respondents was unsure (55%; n=16), and 17% (n=5) indicated that they would probably not be willing to use the tool. The responding IPs were more negative about the attitudes of their colleagues: 90% (n=26) doubted whether their colleagues would be willing to use a decision support tool during work disability assessments. Respondents expected that the tool would be of most benefit for claimants with more complex pathology, such as medically unexplained physical symptoms, in which motivation and perception play an important role. Regarding the question for which types of claimants the tool would be of less benefit, respondents mentioned claimants of which the prognosis is evident (55%) and claimants who lack insight into their own illness (21%). However, 28% of the respondents did not specify specific types of claimants and indicated that the decision support tool could always be consulted. These were not (all) the same respondents as the eight IPs that indicated who they would be willing to use an evidence-based decision support tool based on a prediction model during future work disability assessments.

Barriers and facilitators for use of the decision support tool

The percentages of respondents that agreed, neither agreed nor disagreed, and disagreed that specific barriers or facilitators applied to using of the decision support tool in practice are summarized in figure 1. Among the barriers, wishing to know more about the decision support tool before deciding to apply it showed the highest score (83%). Other barriers were thinking that parts of the decision support tool are incorrect (28%), and that fellow IPs (24%) and other colleagues (21%) would not cooperate in applying the tool. Overall, the mean percentages of IPs who agreed that certain facilitators were applicable to use of the tool were somewhat higher. Concerning the facilitators that were applicable to use of the decision support tool, the majority of the respondents agreed that the tool leaves them enough room to make their own decision (76%), that the layout of the tool makes it handy for use (62%), and that it leaves enough room to weigh the wishes of the claimant (52%).

Table 3. Prognosis without and with the decision support tool in 29 IPs judging 6 case vignettes

	Vignette						
	1 (n=29)	2 (n=29)	3 (n=29)	4 (n=29)	5 (n=29)	6 (n=29)	All (n=174)
Change in WAS predicted by DST[a]	+-	++	+-	--	++	+-	
Without the DST[a]							
Prognosis, n (%)							
Improvement	0 (0)	14 (48)	20 (69)	0 (0)	24 (83)	24 (83)	82 (47)
No change	29 (100)	10 (35)	9 (31)	28 (97)	5 (17)	5 (17)	86 (49)
Deterioration	0 (0)	5 (17)	0 (0)	1 (3)	0 (0)	0 (0)	6 (4)
Concordant prognosis, n(%)	29 (100)	14 (48)	9 (31)	1 (3)	24 (83)	5 (17)	82 (47)
Observed agreement, %	100	48	31	3	83	17	47
With the DST[a]							
Prognosis, n (%)							
Improvement	0 (0)	17 (59)	14 (48)	0 (0)	25 (86)	17 (59)	73 (42)
No change	29 (100)	9 (31)	15 (52)	26 (90)	4 (14)	12 (41)	95 (55)
Deterioration	0 (0)	3 (10)	0 (0)	3 (10)	0 (0)	0 (0)	6 (3)
Changed prognosis, n (%)	0 (0)	3 (10)	6 (21)	2 (7)	1 (3)	7 (24)	19 (11)
Observed agreement, %	100	59	52	10	86	41	58
Change in IPs' prognosis[b]	NS	NS	0.026	NS	NS	0.008	-

[a] DST = decision support tool, ++ = improvement, +- = no change, -- = deterioration

[b] McNemar's test to compare IPs' prognosis before and after use of the decision support tool, *p*-value, NS = not significant ($p>0.05$)

Table 4. Efficacy of a decision support tool on prognosis of work ability and prognostic confidence

	All (n=174)	Concordant prognosis (n=82)	Discordant prognosis (n=92)
Changed prognosis, n (%)	20 (13)	-	20 (22)
Prognostic confidence			
Without the DST[a]	7.1 ± 1.7	7.2 ± 1.5	7.0 ± 1.9
With the DST[a]	6.6 ± 2.2	7.4 ± 1.6	6.0 ± 2.4
ΔNRS[b]	-0.5 ± 2.1*	0.2 ± 1.1	-1.0 ± 2.5*

* Significant decrease in prognostic confidence after using the DST ($p<0.05$)

[a] DST = decision support tool, [b] NRS = numerical rating scale

6

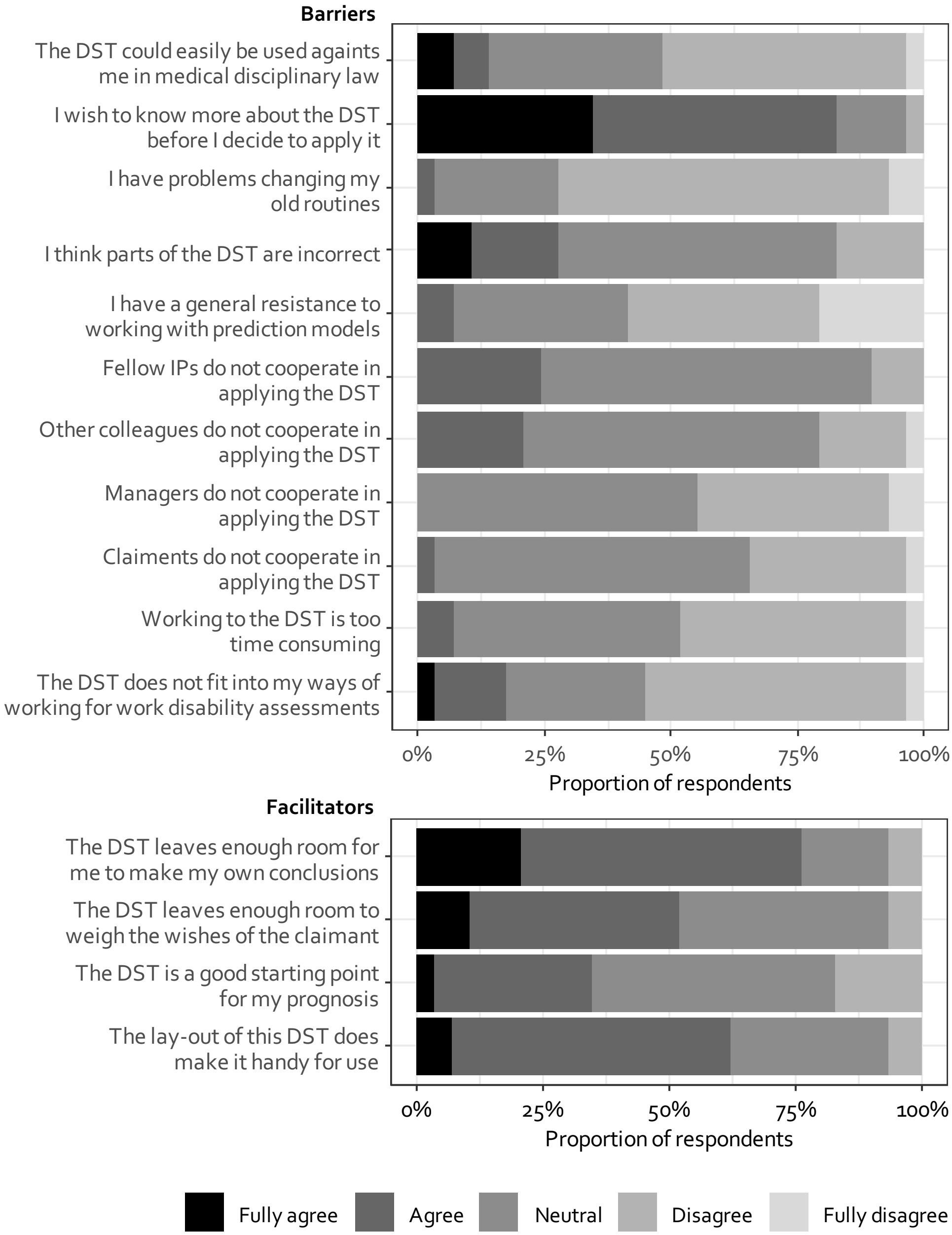


Figure 1. Barriers and facilitators for use of the decision support tool (DST)

Discussion

We found that in only 47% of the cases the prognosis of the IP without information from the decision support tool was concordant with the outcome of the prediction model. In 22% of the cases with a discordant prognosis, IPs changed their prognosis of work ability after consulting the decision support tool. If IP's prognosis was discordant with the outcome of the decision support tool, their prognostic confidence decreased. Although the perceived barriers for use were overall limited, only a minority of the IPs (28%) indicated that they would be willing to use an evidence-based decision support tool based on a prediction model in practice.

In the present study, we found that the agreement between IPs' own prognosis before evaluating the outcome of the decision support tool and the outcome of the prediction model was low. In general, prognosis of future changes in work ability was more positive reported by the IPs than the prognosis estimated by the prediction model. As IPs consider prognosis as the most difficult part of the work disability assessment, this could indicate that IPs often give claimants the benefit of the doubt and, when they are unsure, prefer to plan re-assessments to monitor claimants' change in work ability.

In case IPs' own prognosis was discordant with the outcome of the decision support tool, IPs changed their prognosis in 22% of the cases after they evaluated the tool. This was only 13% of the total number of cases. This low efficacy can possibly be explained by the low initial agreement between IPs' prognosis and the outcome of the decision support tool, which could have influenced IPs' trust in the tool. Although the percentage of changed prognoses in our study was rather low, this is in line with results from a study on the impact of a decision support tool on prediction of progression in early-stage dementia [30]. This prospective multicenter study including 429 patients showed that clinicians changed the prediction of progression only in 13% patients after using the tool. However, the researchers did not mention the number of predictions discordant with the outcome of the decision support tool, and this might have been lower than 53%. A study on Dutch IPs judgement of physical work ability showed that when IPs judgement differed from the information about functional tests, only a one-third changed their judgement [31].

6

IPs' own prognosis was discordant with the outcome of the prediction model in 53% of the cases, but only in 13% of the cases, they changed their prognosis after consulting the tool. In a previous study using retrospective data, we have shown that the prediction model for future changes in work ability can discriminate between claimants with an improvement, deterioration or no change in work ability at one-year follow-up [10]. Internal validation of the prediction model showed that the positive and negative predictive values of the model were 63% and 82%, respectively. Compared to these percentages, the number of cases for which IPs' prognosis was discordant with the outcome of the prediction model, but for which IPs did not change their prognosis (n=73; 42%), was relatively high. This implies that for part of these cases for which the final prognosis of the IP was discordant, the prediction model might have correctly predicted the change in work ability at one-year follow-up. This could be an indication that there is still room for improvement in that IPs could have had more trust in the decision support tool and could have adjusted their prognosis accordingly. However, also in previous studies it was shown that clinicians seem to be reluctant to use decision support tools and other new sources of information in practice [32, 33]. Even if clinicians trust the scientific evidence underlying a decision support tool, they perceive it mostly as a confirmatory tool that can be used to validate or fine-tune their own prognosis rather than substantially change it [12].

High prognostic confidence is important for efficient planning of medical re-assessments and to provide effective interventions for return to work. Mean prognostic confidence without the decision support tool was 7.1 (SD 1.7). For concordant cases, our results did not show a significant change in prognostic confidence. For discordant cases, however, we found a statistically significant decrease of 1 point (SD 2.5). These confidence levels are in line with clinicians' confidence in the prediction of dementia without and with a decision support tool, which ranged between 62-76 (SD 16-19) on a visual analogue scale (0-100%) [34]. In the latter study, they did not discriminate between concordant and discordant cases, and overall a small significant increase in confidence of 3% was found (SD 11). This is in contrast with the significant decrease in confidence of 1 point on the NRS found in the present study.

Transferring effective innovations into real world (medical) settings and achieving sustainable use in every day decision-making processes is a complicated, long-term process [16, 35, 36]. We assessed 15 statements focusing on barriers and facilitators related to knowledge, attitude, and behavior. The majority of the respondents indicated that they currently doubted whether they would use the decision support tool for prognosis of work ability during the work disability assessment. Overall, the reported barriers among our IPs were limited. The main perceived barrier was that IPs felt that they needed more information about the decision support tool before they could decide whether to use the tool in practice. Lack of knowledge was also mentioned as one of the main barriers for adherence to guideline recommendations in a study among Dutch general practitioners [37]. The need for more information could have influenced the perceived usefulness of the decision support tool, which is a key characteristic for acceptance and use [38, 39]. In the present study, this need is possibly reinforced by the low agreement between the outcome of the prediction model and IPs' prognosis of future work ability, and the fact that IPs first had to estimate their own prognosis while the decision support tool was only presented after their decision-making process.

The most important facilitators for use of the decision support tool were as follows: the majority of IPs believed that the decision support tool leaves them enough room to make their own conclusions (80%) and that the layout of the tool was perceived as handy for use (67%). The former may have arisen from the fact that in the present study IPs were first asked to assess their own prognosis, and subsequently they were informed about the outcome of the decision support tool. The latter seems a favorable factor for future use of the tool, as clear graphical representation supports the understanding of the decision support tool and leads to better informed decision-making [40, 41].

Strengths and limitations

The main strength of our study is that we focused on the efficacy of and attitudes towards actual use of an evidence-based decision support tool that was developed for and in accordance with Dutch IPs. Moreover, because of the within-subject design, respondents could act as their own control group. For each case vignette, respondents were first asked to give their prognosis based on the case description, after which the decision support tool with the outcome of the prediction model was presented. The former prognosis and prognostic confidence can be considered as control data, and compared with the outcome measures after the respondents consulted the tool.

Another strength of the study was that the participants were recruited out of all potential users of the decision support tool. There are small differences in working procedures between offices, as well as differences between claimants' populations. As IPs from seven different offices participated in this study, we were able to gather different perspectives on the benefits of the tool, and on possible barriers and facilitators for use. The number of IPs that voluntary wanted to participate in the study (n=29) was higher than the minimum number of IPs. However, a potential selection bias was introduced, as participation was voluntary. This may have led to an increased sampling of like-minded IPs who are interested in prediction models and the use of evidence-based decision support tools in practice.

A second limitation is that, for practical and ethical reasons, we were unable to perform the randomized controlled trial (RCT) that we originally planned and could therefore not externally validate the model in a new dataset. Instead, we had to rely on evaluation of the decision support tool in an experimental setting. Using case vignettes, we could not assess the prognostic accuracy of either the decision support tool or the prognosis of the IPs. Although the case vignettes were based on real work disability assessment reports, they were fictitious descriptions of claimants and tested in an experimental setting. Hence, this did not allow us to make any statement about the level of work ability at one-year follow-up. Moreover, performing an experimental study as a substitute for a RCT meant that we could not consider any learning effects. In general, users need some time to get comfortable with a change. As the use of predictive analytics and decision support tools in insurance medicine is rather limited, it could be expected that their impact will gradually increase over time.

Implications for research and practice

The use of evidence-based decision support tools in occupational health could improve decision-making processes and quality of work disability assessments. Such tools can guide IPs when they are unsure about the prognosis, and unveil prognostic factors that are important to draw attention to in the return to work process. In particular, a decision support tool has the potential of a supportive effect when prognosis of work ability is unclear. The tool was considered to be of less benefit for claimants of which the prognosis is evident. In the present study, IPs first made their own prognosis and subsequently evaluated the decision support tool. Although this is in line with the results of a previous study, in which professionals indicated that they would first want to make their own prognosis and afterwards verify or adjust their evaluation based on the outcome of decision support tool, this might have lowered the efficacy of the tool [11]. Presenting the tool at the beginning of the decision-making process might increase its efficacy by making the tool part of this process. Once external validation of the prediction model has confirmed its accuracy, it could be given the same value as other sources of information that IPs use, such as medical guidelines, information from treating physicians and the medical history of the claimant.

Effectiveness of implementation processes is strongly associated with innovations being carefully implemented and free from serious implementation problems. Insight in the perceived barriers and facilitators for use in work disability assessments can be used to design the future implementation process of the decision support tool. Only a minority of the IPs that participated in this study indicated that they would use a decision support tool in practice. Unfamiliarity with prediction models and decision support tool was mentioned as the main barrier that may prohibit IPs from using such tools. At the start of the present study, we organized short meetings to give IPs more information about the prediction model and decision support tool, and to inform them about the design and questionnaires of the experimental study. However, these short meetings might not have been sufficient. Participants might wish to have more detailed information on the prediction model, for instance on the weights of the prognostic factors that were included in the regression equation. To improve knowledge, it may be useful to conduct more extensive information sessions on the use of prediction models and decision support tool in insurance medicine or incorporate these topics in training programs. Moreover, to build trust with future users, the decisions support tool should be validated in occupational health practice. Therefore, future research should pay attention to the effect of using the decision support tool in everyday practice, focusing on IPs prognosis of work ability and prognostic confidence in work disability claimants with different types of diagnoses.

Although individual participants were rather reluctant towards use of a decision support tool in real practice, it may be beneficial from an organizational perspective. IPs' prognosis during the initial work disability assessment of a claimant is used to schedule subsequent re-assessments. Our results indicate that IPs' assessment of prognosis is relatively positive. Given the limited capacity of occupational health resources, and especially a shortage in the number of IPs at the SSI, the SSI might want to minimize the re-assessments for claimants for which no relevant change in work ability is expected. A prediction model for prognosis can help to use the limited available capacity as effective as possible, i.e. to plan re-assessments for claimants who will benefit most from contact with an IP. In this regard, elaborate knowledge transmission and embedding the tool in working policies seem important factors. Using decision support tools at an organizational level involves some ethical and medicolegal considerations for professionals and policy makers. Hence, it should be emphasized that decision support tools are not meant as stand-alone or management tools, but as a complement to professionals' own estimates of prognosis.

The findings of the present study could be used when developing other prediction models and decision support tools for occupational health professionals. Our tool focuses on prognosis of work ability for claimants who were sick-listed for two years. However, the longer individuals are absent from work, the less likely they are to return [42-44]. A prediction model for workers who have just recently been sick-listed can help occupational health professionals to target individuals at risk of long-term sickness absence and identify effective early interventions [45, 46]. From a rehabilitation point of view, such a tool could have more impact. Our findings on the attitudes of IPs towards prediction models and decision support tools include general aspects that could be helpful when developing such a tool.

Concluding remarks

The present study showed that the congruence of the decision support tool with IPs' prognosis of future work ability was low, and that IPs' prognostic confidence decreased after evaluating the tool if their prognosis was discordant with the outcome of the prediction model. Only a minority of the IPs changed their prognosis when it was discordant with the outcome of the tool. Most IPs indicated that they were unsure or they were not willing to use an evidence-based decision support tool based on a prediction model during future work disability benefits. Unfamiliarity with prediction models and decision support tool was mentioned as the main barrier that may prohibit IPs from using such tools.

Appendix

Example of a case vignette used in the study

Description of the sixth case vignette that was given to IPs for consideration:

- Until two years ago, Yvonne (45 years old) worked as an administrative assistant for 40 hours a week. Her job had a heavy workload, and required high levels of concentration and multitasking.
- She had to call in sick at work after she had a collision while riding her bike. This accident caused multiple fractures on both arms and knee complaints. Her physiatrist indicated that Yvonne has humerus fractures on her left and right arm, and that she got a surgery with metal screws and plates. She follows a rehabilitation trajectory, but due to her mental disorders, this runs longer than expected.
- A letter from her treating psychologist states that Yvonne suffers from a post-traumatic stress disorder and a mood disorder. When she was a child, she suffered from child traumatic stress. This caused instability in her mood and personality development. Yvonne has regular consults with her psychologist, but so far, she rejects a medical treatment.
- During the work disability assessment interview, Yvonne looks vulnerable. She says that she experiences many mental and physical limitations in daily functioning. She thinks she is not yet ready to return to work, but expects that her limitations will reduce in the future.
- Yvonne has limitations related to personal and social functioning: she has difficulty concentrating, often experiences sensory overload, avoids conflicts, and is very sensitive to other people's feelings. Due to her physical complaints, she cannot optimally use her musculoskeletal system without pain and has difficulties with dynamic movements (lifting, twisting, bowing, walking at work etc.). On medical grounds, she gets a restriction in working hours for maximum 4 hours a day, 20 hours a week.
- Yvonne has psychological and orthopedic treatment. Three times a week she has an appointment with her physiotherapist. It is expected that her medical conditions and functional possibilities will substantially improve in the future.

Prognostic factors that were taken into account in the prediction model:

- If you would give 10 points to your work ability during the best period in your life, how many points (between 0-10) would you give to your level of work ability of the past 2 weeks? *0 points*
- During the past 4 weeks, due to your physical health, did you achieve less than you would in your work other daily activities? *Yes*

- How often during the past 4 weeks, did you feel so down that nothing could cheer you up? *Often*
- How often during the past 4 weeks, did you feel calm and peaceful? *Seldom*
- Since the first day of sickness absence, how often have you visited a treating practitioner? *More than 10 times*
- Type of comorbidity: comorbidity of physical and mental disorders
- Expected change in work ability at one-year follow-up as predicted by the prediction model: *no expected change in work ability*

References

1. De Boer W-E, Besseling J, Willems J. Organisation of disability evaluation in 15 countries. Prat Organs Soins. 2007;3(38):205-217.
2. Kok R, Hoving JL, Verbeek J, Schaafsma FG, van Dijk FJ. Integrating evidence in disability evaluation by social insurance physicians. Scand J Work Environ Health. 2011;37(6):494-501.
3. Anner J, Schwegler U, Kunz R, Trezzini B, de Boer W. Evaluation of work disability and the international classification of functioning, disability and health: what to expect and what not. BMC Public Health. 2012;12(1):470.
4. Kok R, Verbeek JA, Faber B, van Dijk FJ, Hoving JL. A search strategy to identify studies on the prognosis of work disability: a diagnostic test framework. BMJ open. 2015;5:e006315.
5. Kox RJ, Hoving JL, Verbeek JH, Schouten MJ, Hulshof CT, Wind H, et al. Assessment of prognosis by physicians involved in work disability evaluation: A qualitative study. PloS One. 2019;14(2):e0212276.
6. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ. 2005;330(7494):765.
7. Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, et al. Clinical decision support systems for the practice of evidence-based medicine. J Am Med Inform Assoc. 2001;8(6):527-534.
8. Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: a meta-analysis. Psychol Assess. 2000;12(1):19-30.
9. Aegisdottir S, White MJ, Spengler PM, Maugherman AS, Anderson LA, Cook RS, et al. The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. J Couns Psychol. 2006;34(3):341-382.
10. Louwerse I, Huysmans MA, van Rijssen JH, Schaafsma FG, Weerdesteijn KH, van der Beek AJ, et al. Predicting future changes in the work ability of individuals receiving a work disability benefit: weighted analysis of longitudinal data. Scand J Work Environ Health. 2020;46(2):168-176.
11. Louwerse I, Huysmans MA, van Rijssen JH, Overvliet J, van der Beek AJ, Anema JR. Preferences regarding the way of use and design of a work ability prognosis support tool: a focus group study among professionals. Disabil Rehabil. 2019; doi:10.1080/09638288.2019.1693643.
12. Hallen SA, Hootsmans NA, Blaisdell L, Gutheil CM, Han PK. Physicians' perceptions of the value of prognostic models: the benefits and risks of prognostic confidence. Health Expect. 2015;18(6):2266-2277.
13. Van Muijen P, Duijts S, Kornet-Van der Aa D, Van der Beek A, Anema J. Work disability assessment of cancer survivors: insurance physicians' perspectives. Occup Med. 2015;65(7):558-563.
14. Lugtenberg M, Burgers JS, Besters CF, Han D, Westert GP. Perceived barriers to guideline adherence: a survey among general practitioners. BMC Fam Pract. 2011;12(1):98.
15. Dekkers-Sánchez PM, Wind H, Frings-Dresen MH, Sluiter JK. Implementation of a checklist to assess factors relevant for work ability assessments of employees on long-term sick leave. Int Arch Occup Environ Health. 2015;88(5):577-588.
16. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. Lancet. 2003;362(9391):1225-1230.
17. Grol R. Beliefs and evidence in changing clinical practice. BMJ. 1997;315(7105):418-421.
18. Anner J, Kunz R, Boer Wd. Reporting about disability evaluation in European countries. Disabil Rehabil. 2014;36(10):848-854.

19. Spanjer J, Krol B, Popping R, Groothoff JW, Brouwer S. Disability assessment interview: the role of detailed information on functioning in addition to medical history-taking. J Rehabil Med. 2009;41(4):267-272.
20. Schellekens J, Abma F, Mulders H, Brouwer S. Measuring clients' perception of functional limitations using the Perceived Functioning & Health questionnaire. J Occup Rehabil. 2010;20(4):512-525.
21. Weerdesteijn KH, Schaafsma FG, Louwerse I, Huysmans MA, Van der Beek AJ, Anema JR. Does self-perceived health correlate with physician-assessed functional limitations in medical work disability assessments? J Psychosom Res. 2019;125:109792.
22. Ilmarinen J. Work ability—a comprehensive concept for occupational health research and prevention. Scand J Work Environ Health. 2009;35(1):1-5.
23. Ahlstrom L, Grimby-Ekman A, Hagberg M, Dellve L. The work ability index and single-item question: associations with sick leave, symptoms, and health–a prospective study of women on long-term sick leave. Scand J Work Environ Health. 2010;36(5):404-412.
24. Boström M, Sluiter JK, Hagberg M. Changes in work situation and work ability in young female and male workers. A prospective cohort study. BMC Public Health. 2012;12(1):694.
25. Wallander L. 25 years of factorial surveys in sociology: A review. Soc Sci Res. 2009;38(3):505-520.
26. Rossi PH, Anderson AB. The factorial survey approach: An introduction, in Measuring social judgments: The factorial survey approach, Rossi PNock S, Editors. 1982, Sage: Beverly Hills, CA.
27. Atzmüller C, Steiner PM. Experimental vignette studies in survey research. Methodology. 2010;6(3):128-138.
28. Hughes R, Huby M. The application of vignettes in social and nursing research. J Adv Nurs. 2002;37(4):382-386.
29. Peters M, Harmsen M, Laurent M, Wensing M. Ruimte voor verandering? Knelpunten en mogelijkheden voor verandering in de patiëntenzorg [Room for improvement? Barriers to and facilitators for improvement of patient care], in IQ healthcare Scientific Institute for Quality of Healthcare. Radboud University Medical Center Nijmegen. 2002, Centre for Quality of Care Research (WOK), Radboud University Nijmegen Medical Centre: Nijmegen.
30. Bruun M, Frederiksen KS, Rhodius-Meester HF, Baroni M, Gjerum L, Koikkalainen J, et al. Impact of a clinical decision support tool on prediction of progression in early-stage dementia: a prospective validation study. Alzheimers Res Ther. 2019; doi:10.1186/s13195-019-0482-3.
31. Wind H, Gouttebarge V, Kuijer PPF, Sluiter JK, Frings-Dresen MH. Effect of functional capacity evaluation information on the judgment of physicians about physical work ability in the context of disability claims. Int Arch Occup Environ Health. 2009;82(9):1087-1096.
32. Linder J, Schnipper J, Tsurikova R, Yu T, Volk L, Melnikas A, et al. Documentation-based clinical decision support to improve antibiotic prescribing for acute respiratory infections in primary care: a cluster randomised controlled trial. J Innov Health Inform. 2009;17(4):231-240.
33. Rubin MA, Bateman K, Donnelly S, Stoddard GJ, Stevenson K, Gardner RM, et al. Use of a personal digital assistant for managing antibiotic prescribing for outpatient respiratory tract infections in rural communities. J Am Med Inform Assoc. 2006;13(6):627-634.
34. Bruun M, Frederiksen KS, Rhodius-Meester HF, Baroni M, Gjerum L, Koikkalainen J, et al. Impact of a clinical decision support tool on dementia diagnostics in memory clinics: The PredictND validation study. Curr Alzheimer Res. 2019;16(2):91-101.
35. Durlak JA, DuPre EP. Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. Am J Community Psychol 2008;41(3-4):327-350.

36. Powell BJ, McMillen JC, Proctor EK, Carpenter CR, Griffey RT, Bunger AC, et al. A compilation of strategies for implementing clinical innovations in health and mental health. Med Care Res Rev. 2012;69(2):123-157.
37. Lugtenberg M, Zegers-van Schaick JM, Westert GP, Burgers JS. Why don't physicians adhere to guideline recommendations in practice? An analysis of barriers among Dutch general practitioners. Implement Sci 2009;4(1):54.
38. McDermott L, Yardley L, Little P, Ashworth M, Gulliford M. Developing a computer delivered, theory based intervention for guideline implementation in general practice. BMC Fam Pract. 2010;11(90).
39. Kortteisto T, Komulainen J, Mäkelä M, Kunnamo I, Kaila M. Clinical decision support must be useful, functional is not enough: a qualitative study of computer-based clinical decision support in primary care. BMC Health Serv Res. 2012;12(1):349.
40. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. Psychol Sci Public Interest. 2007;8(2):53-96.
41. Trevena LJ, Zikmund-Fisher BJ, Edwards A, Gaissmaier W, Galesic M, Han PK, et al. Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. BMC Med Inform Decis Mak. 2013;13(2):S7.
42. Schuring M, Robroek SJ, Otten FW, Arts CH, Burdorf A. The effect of ill health and socioeconomic status on labor force exit and re-employment: a prospective study with ten years follow-up in the Netherlands. Scand J Work Environ Health. 2013:134-143.
43. Henderson M, Harvey SB, Øverland S, Mykletun A, Hotopf M. Work and common psychiatric disorders. J R Soc Med. 2011;104(5):198-207.
44. Lännerström L, Wallman T, Holmström IK. Losing independence–the lived experience of being long-term sick-listed. BMC Public Health. 2013;13(1):745.
45. Nieuwenhuijsen K, Verbeek JH, de Boer AG, Blonk RW, van Dijk FJ. Predicting the duration of sickness absence for patients with common mental disorders in occupational health care. Scand J Work Environ Health. 2006;32(1):67-74.
46. Louwerse I, van Rijssen HJ, Huysmans MA, van der Beek AJ, Anema JR. Predicting Long-Term Sickness Absence and Identifying Subgroups Among Individuals Without an Employment Contract. J Occup Rehab. 2020:1-10.

CHAPTER 7

General discussion

The main aim of this thesis was to investigate how data analytics, prediction models and decision support tools can help insurance physicians (IPs) in the Netherlands to make evidence-based decisions regarding prognosis of functional abilities and support to return to work. The current chapter starts with a summary of the main findings of this thesis and some methodological considerations. In addition, the potential use of prediction models in insurance medicine, and the challenges and considerations for implementation are discussed. Finally, this chapter ends with recommendations for future research, and implications for policy and practice.

Main findings

Factors associated with work disability entitlement and duration

In chapter 2, a cohort study of 31,733 individuals receiving work disability benefits from the Dutch Social Security Institute (SSI) showed that mental disorders were the most frequent diagnosis for individuals claiming long-term work disability. Diagnoses differed among age groups and education categories; whereas mental disorders were the main diagnosis for work disability among younger and more highly educated individuals, physical disorders (generally musculoskeletal, cardiovascular and cancer) were the main diagnosis among older and less educated individuals.

Considering a 5-year follow-up period, we found that in 82% of the claims the duration of disability benefit was five years or more after approval. Especially individuals with (multiple) mental disorders, and those with comorbidity of mental and physical disorders were at high risk for continuing eligibility for disability benefits. These results indicate that return to work of claimants receiving a work disability benefit is challenging. This emphasizes that return to work support for sick listed individuals and work disability benefit recipients is important.

Risk of long-term sickness absence

Early identification of individuals at risk of long-term sickness absence and work disability can help occupational health professionals to target specific at-risk groups and offer them effective interventions for return to work. In chapter 3, we conducted a longitudinal study on prognostic factors for long-term sickness absence among sick listed workers without an employment contract. We showed that almost one-third of the study population was still at sickness absence one year after approval of the benefit. Based on three variables (educational level, self-reported expected sickness absence duration, and self-reported help-seeking ability), we could fairly discriminate between individuals with and without long-term sickness absence. Using the predicted risk of long-term sickness absence in combination with self-reported variables, four subgroups were identified: sick listed workers with mental limitations, sick listed workers with physical limitations, sick listed workers with positive expectations, and sick listed workers with negative expectations about recovery and return to work. This allocation of workers in different subgroups could contribute to efficient allocation of return to work interventions tailored to the groups that would benefit most from it.

A prediction model and decision support tool for future changes in work ability

Estimating prognosis of work disability is one of the main tasks of IPs during the work disability assessment. Because of the limited occupational health resources, accurate prognosis of work ability is important to identify those in need of return to work interventions and for efficient planning of medical re-assessments. To assist IPs to make an accurate prognosis of work disability, we developed a model that predicts future changes in work ability (chapter 4) using data from a prospective cohort study. Once a work disability benefit was granted, the majority of claimants (63%) did not experience a change in work ability one year after approval of the benefit. Work ability at baseline was the strongest predictor for a change in work ability one year after approval of the work disability benefit. Moreover, several mental and physical functioning factors, work status, and wage loss were found as prognostic factors. For instance, claimants with a wage loss of more than 80% had a smaller probability of experiencing an improvement in work ability.

The prediction model, which was developed using the data of this prospective cohort of work disability claimants, generates claimant-specific predictions. Based on this prediction model, a decision support tool was developed that uses characteristics of individual claimants as input. In order to develop a useful and relevant tool, we conducted focus groups to gain insight in how and where in the decision-making process professionals would like to use the tool. Professionals mentioned that, to increase the chance that they will use the tool, it should be easy to access and interpret (chapter 5). In addition, they did not want to know the outcome of the tool before the assessment to prevent bias in their decision-making. Concerning preferences on the design of the tool, dividing work disability claimants into categories based on the outcome of the prediction model and assigning colors to these categories was experienced as the most straightforward and clear way of presenting the results.

Efficacy and feasibility of the decision support tool

A vignette study showed that the efficacy of the tool on IPs' prognosis of work ability and their prognostic confidence was moderate, i.e. only a minority of IPs changed their prognosis after evaluating the outcome of the prediction model (chapter 6). If IPs' prognosis was not in line with the outcome of the prediction model, their prognostic confidence decreased. There was no change in prognostic confidence for cases where the prognosis of the IP was concordant with the outcome of the prediction model. Concerning IPs attitudes towards use of the tool, the main barrier was that IPs wanted to know more about the tool before they would decide to apply it. Although the perceived barriers were overall limited, only a minority of the IPs indicated that they would be willing to use the tool in practice.

Methodological considerations

In this section some methodological considerations regarding data quality, data availability and generalizability of the findings of this thesis are discussed.

Data quality

In this thesis, datasets of three cohorts were used to gain insight in prognostic factors for long-term sickness absence and work (dis)ability. These cohorts were merely based on registration data of the SSI (chapter 2), or combined these registration data with self-reported questionnaires filled in by claimants at the moment they applied for a sickness absence (chapter 3) or work disability benefit (chapter 4). The advantage of using registration data is that these data are collected anyhow and hence no additional data collection processes need to be set up. In this way, using registration data might save research resources, and it does not pose an unnecessary burden on claimants and professionals by asking them information that is already available. However, an overall concern in our studies is that the registration data of the SSI were not collected for research purposes, but instead were registered by SSI employees for administrative reasons. Although accurate and correct registrations are important for working processes at the SSI, professionals are unaware of research implications of incorrect records. This implies that the data might have been treated differently than when collected for research purposes. For instance, data were not collected with validated instruments and reported in a structured way, i.e. the absence of findings in records of a disability assessment is usually not registered, and records can contain registration errors. Hence, careful interpretation and processing of registration data is needed before these can be used in research. Next to concerns about data quality, another disadvantage of using registration data is that researchers have no control over the variables for which data is collected. Some of the variables that, based on previous studies, are expected to be relevant prognostic factors, may be perceived by IPs to be irrelevant for practical purposes. If for that reason they are not included in the registration data, then they are also not available for research.

In chapters 3 and 4, we combined registration data with self-reported questionnaires. The questionnaire data of chapter 3 were collected as part of a pilot study within the SSI. For chapter 4, the self-reported data were retrieved from the Forward study [1]. As shown in figure 1, the potential prognostic factors that are registered in the registration data of the SSI are different from those collected using the self-reported validated questionnaires. Combing these two data sources enabled us to explore a broad range of sociodemographic, health, work, and psychological factors. As for the quality of registration data, there are also some concerns about the quality of self-reported data. The main concern of using self-reported measures in behavioral and medical research is response bias [2]. Among the many potential reasons for offering biased estimates of self-assessed measures are misunderstanding of the question and socially desirable responding [3]. In chapter 3, we only used self-reported measures at baseline as prognostic factors for registered sickness absence at one-year follow-up. In chapter 4, we also used the self-reported change in work ability between baseline and one-year follow-up as dependent variable in the prediction model. Studies have shown that there can be a high variation in the level of work ability that physicians and claimants report [4, 5]. In general, claimants rate their work ability lower than physicians do [6, 7]. The correlation between claimants' and physicians' assessments of work ability are especially low for claimants with somatoform and depressive disorders [1, 5, 8]. However, as we were not interested in the level of work ability in itself, but in the change in work ability, this discrepancy between physicians' and claimants' assessments has been of less influence in our study. This assumption is confirmed by the finding that claimants and physicians seem to predict improvement of functional limitations after the benefit has been granted with about the same level of accuracy [9, 10]. Moreover, response bias in cohort studies seems to be especially problematic when individuals are subject to an instrumentation-related source of contamination, known as response-shift bias [11]. This would, for instance, be the case when individuals are exposed to an intervention that changes their perception of the concept of work ability, without necessarily changing their work ability itself. In the cohort study of chapter 4, the respondents just received the normal return to work support of the SSI, and were not exposed to any study-related treatment or interventions. Therefore, we expect that respondents' understanding and

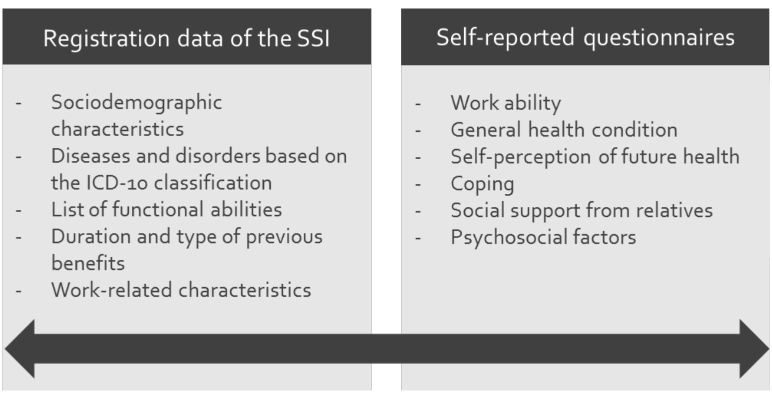


Figure 1. Combining two data sources

interpretation of the concept of work ability remained unchanged, and the impact of response-shift bias on our findings could be neglected.

Another aspect concerning self-reported data is its trustworthiness. This is particularly the case when the questions concern private or sensitive topics, such as self-report of health and personal factors. Due to the large number of rules and regulations that apply to the Dutch social security system, the process of applying for a work disability benefit and the rules for benefit entitlement can be unclear for claimants. As not being granted a work disability benefit can have huge personal and financial consequences, claimants are likely to experience feelings of uncertainty during the process of benefit application [12, 13]. Participants of our focus group study (chapter 5) and vignette study (chapter 6) argued that claimants who are aware of the factors that contribute to the predicted change in work ability could provide biased answers. However, as we combined self-reported data with registration data of social security institutes, trustworthiness issues can be expected to be of minor importance.

Data availability

Additionally, relying mainly on existing data from cohorts resulted in some limitations regarding the statistical analyses and practical implications. One of the restrictions concerned the relatively low sample sizes. The study populations of chapters 3 and 4 consisted of 437 and 944 individuals, respectively. This posed some restrictions on the statistical analyses that we could apply, as more advanced statistical techniques generally require a larger number of observations than traditional regression models. Machine learning models have high potential as they can handle enormous numbers of prognostic factors and combine them in nonlinear and interactive ways, but they also require thousands of observations to reach acceptable performance levels [14]. The relatively small sample sizes of our study populations also limited possibilities for internal validation of our prediction models. In chapter 4, we used 20% of the study population as a hold-our test set to validate the prediction model. The resulting test set contained a sufficient number of 187 claimants, but larger sample sizes would have allowed more rigorous conclusions [15]. Finally, we had to deal with gaps in the data resulting from working processes and, as we used existing data from cohorts that were collected to answer other research questions, with the in- and exclusion criteria that the researchers collecting data for these cohorts originally formulated. Because of the former, claimants with a full and permanent work disability benefit had to be excluded from part of the analyses in chapter 2, as the SSI did not register their educational level. Concerning the latter, we were not able to predict future changes in work ability of individuals diagnosed with cancer or individuals suffering from psychotic disease or dementia as they were excluded from participation in the Forward study (chapter 4) [1]. In addition, we were unable to use unstructured data from work disability assessments as these were not recorded or collected in a structured way. Unstructured data include, for instance, data embedded in occupational health reports or recordings from the work disability assessment interview with a claimant. These data can offer a wealth of information relevant to understanding sickness absence and work disability, but we were unable able to take them into account in our studies [16].

Generalizability

Finally, the findings of this thesis should be considered in the context of the Dutch social security system. The study samples of this thesis are fairly representative for the Dutch populations of long-term sickness absence and work disability claimants. However, worldwide there are large differences between disability insurance systems, for instance in terms of workers' compensation, the role of different professionals in disability assessment and monitoring, and in support to return to work processes [17, 18]. In most countries, eligibility for work disability benefits is based on a medical certificate, which is provided by the treating physician or general practitioner. In contrast, in the Netherlands the regular healthcare system does not play an important role in occupational health services [19]. Moreover, whereas in the Netherlands sickness absence and work disability benefits are provided regardless of the cause of the disease or disability, in most other countries there is a distinction between occupational causes and non-occupational causes. Hence, our study population is not representative for populations of claimants in other social security systems, for instance in terms of reason of sickness absence or continuing eligibility for benefits. These differences should be taken into account when interpreting our study findings and translating them to other systems.

Perspectives on the impact of data analytics and prediction models

The use of data analytics and prediction models is more common in clinical practice than in insurance medicine. This thesis shows that they can also be of value for IPs and other occupational health professionals. First, this section discusses the use and value of prediction models in clinical practice. Next, the potential benefits of prediction models for insurance medicine will be explored.

Using data analytics in clinical practice

In clinical practice, prediction models combine patient characteristics, test results, and/or other disease characteristics to predict medical outcomes. These can either be predictions regarding the probability of the presence of a disease (diagnosis) or an event in the future course of the disease (prognosis). Many studies have shown the possibility to transform data into valuable knowledge for clinical practice. For instance, some recent publications have shown that prediction models in clinical practice might be able to diagnose different types of cancer in real-time, and with at least the same accuracy as clinicians. More specifically, these publications reported on: (1) a model that predicts brain tumor diagnosis in the operating room with the same accuracy as pathologists and much faster than conventional techniques [20]; (2) a model that not only identifies prostate cancer from biopsies, but can also grade the tumor at a performance level similar to that of specialized urological pathologists [21]; and (3) a new model for breast cancer prediction, based on two large clinically representative datasets, that is capable of surpassing six radiologists in breast cancer prediction [22].

These examples show that prediction models could improve diagnostic accuracy. They can advise clinicians in the medical decision-making process. By supporting professionals' decision-making, prediction models can result in increased quality of the time with the patient, better overall quality of services by the avoidance of errors, and improved efficiency. For instance, double reading and screening of mammograms is recognized as the best method for the detection of small invasive cancers, but this is often difficult to achieve in practice due to high costs and the need for two radiologists [23]. With the accuracy of cancer detection rates of a single radiologist using computer-aided detection being similar to those of two radiologists, the use of clinical decision support tools can overcome these barriers [24]. Next to the possible direct impact on diagnoses and prognoses, clinical decision support tools also benefit clinicians by giving a different perspective, providing feedback and enhancing critical thinking [25, 26].

Potential benefits of prediction models for insurance medicine

In insurance medicine, prediction models could be useful for prognosis of work disability of claimants and for selecting rehabilitation interventions. We stress that prediction models are not meant to take over the job of IPs, likewise they are also not taking over the job in clinical practice, but to support IPs to make more evidence-based decisions by providing objective estimates of outcome probabilities to complement IPs' expertise and existing information resources. The latter include, for instance, the medical history of a claimant, information from treating physicians, and insurance medicine guidelines. A decision support tool could be added as an additional source of information that can assist IPs in the decision-making process.

Just like it is for clinical practice, prediction of prognosis is also relevant in the assessment of work disability [27]. In insurance medicine, prognosis means estimating either improvement, stabilization or deterioration of a claimant's functional limitations and work ability [28]. In the Netherlands, IPs should plan and conduct medical re-assessments to monitor changes in functional limitations and work ability. Due to a shortage of IPs, however, there is limited capacity to perform these re-assessments. At the same time, IPs do not always assess a relevant change in functional limitations and work ability when they conduct re-assessments [29]. This is a missed opportunity as they could have performed a re-assessment for another claimant instead. It is known that for claimants with an improvement in work ability timely re-assessments have a positive effect on return to work [30, 31]. In chapter 4, we have shown that only a minority of the claimants (22%) experience an improvement in work ability at one-year follow-up and that the prediction model performs well in predicting cases that will not experience a change in work ability. Hence, the use of this prediction model in order to get to know which claimant should not be assessed at a particular moment, because a change in work ability is unlikely, would help effective allocation of resources. Moreover, in chapter 6, the results of the efficacy study showed that in 53% of the cases IPs' own prognosis of work ability was not in line with the outcome of the prediction model, and that in most of these cases IPs' prognosis of work ability was more positive (88%) compared to the prediction model. A prediction model for prognosis can help to use the limited available capacity as effective as possible, i.e. to plan re-assessments for those claimants that will benefit most from follow-up contact with an IP.

Furthermore, decision support tools can improve understanding of the factors that are related to long-term sickness absence or future changes in work ability [32]. Next to IPs, other occupational health professionals, such as labour experts or occupational physicians, can benefit from data analytics and prediction models as well. Rehabilitation professionals guide claimants in the return to work process. Because resources are limited, these professionals need to decide to whom, when, and which return to work intervention should be offered. The potential effectiveness of interventions depends on several characteristics of a claimant, such as health-related and personal factors [33]. Therefore, a differentiated approach is needed. A division of claimants in distinct subgroups, based on several personal characteristics and the predicted risk of long-term sickness, as presented in chapter 3, can be helpful. This division in subgroups can guide rehabilitation professionals in choosing for which claimants simple, low-intensity and low-cost interventions might be adequate, and for which claimants more intensive and structured interventions might be required. In a similar way, professionals can use more accurate predictions of future changes in work ability to decide, in consultation with a claimant, about the optimal timing and type of return to work intervention. In this manner, data analytics can assist in ensuring that return to work interventions are allocated in an efficient and appropriate manner, such that they meet individual claimants needs. Offering a personalized approach prevents that claimants have to put effort in ineffective interventions, and can save valuable time and money from an organizational perspective.

Use of prediction models could also be seen as a way to achieve higher inter-rater reliability in work disability assessments. Social security laws and regulations prescribe that individuals with similar health injuries and exposed to similar working conditions should receive similar judgements of medical impairments and functional limitations. Despite huge impact on individuals' personal life and rehabilitation, studies have revealed systematic variations in work disability assessments among professionals [34]. Reasons for differences in assessment of the same claimant include information variance (asking different questions and thereby obtaining different information), observation variance (differences in what professionals notice and remember), interpretation variance (differences in the significance attached to what is observed), and expert variance (differences in understanding of job demands and the level of effort that can be expected of a claimant) [35, 36]. Due to lack of time and difficulties in translating evidence to the situation of an individual claimant, IPs mainly rely on their own competencies and experience when estimating prognosis of work ability [37]. The systematic variation in disability assessments and prognosis can be reduced by using well-developed instruments and guidelines that standardize the collection, interpretation and reporting of information [38]. Using a clinical decision support tool, as auxiliary source of information, could reduce professionals' uncertainty and further lower the degree of variation among professionals [39].

How to overcome professionals' objections to using prediction models

Although many prediction models have been developed in scientific studies, only few of them are actually being used in practice to support clinical decision-making [40, 41]. Physicians acknowledge the benefits of prediction models, but also perceived these models as having potential risks, which hinder their use [42]. Below, we discuss some of the perceptions and

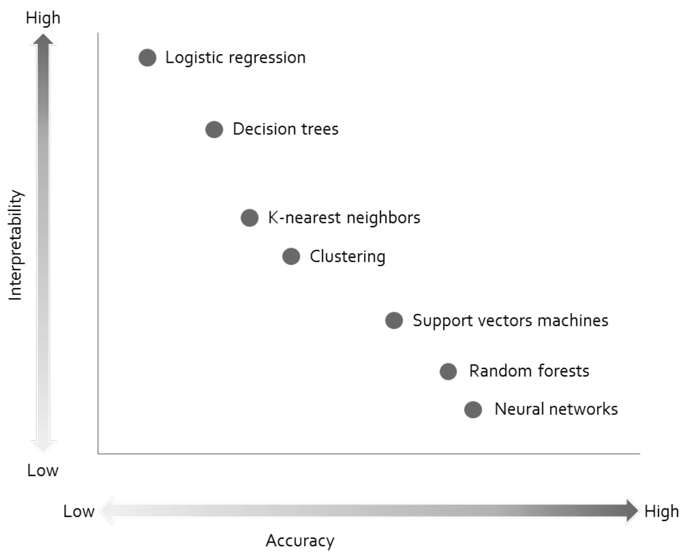


Figure 2. Interpretability-accuracy trade-off in prediction models

beliefs that currently hinder occupational health professionals to use data analytics and predictions models up to their full potential.

The balance between interpretability and accuracy

A comment that was frequently made by IPs and other SSI professionals in our focus group study (chapter 5) as well as in other meetings and presentations about our research was that full understanding of the prediction model is a prerequisite to be able to trust the outcome of the model. This posed restrictions on the statistical analyses that we could apply, as not all types of prediction models can equally well be explained. Although comprehensive oral and written information was given prior to the start of the experimental study, the results of chapter 6 show that the main barrier for use of the decision support tool in practice was still related to the need for more information. In this study, the majority of the respondents (83%) indicated that they wished to know more about the content of the tool before they would decide to apply it in practice. Professionals indicated that they want to have complete knowledge of the prediction rule, how it was developed, its prognostic factors, and of the uncertainty around the predicted outcomes. For instance, they wanted to know the contribution of each prognostic factor in order to decide whether they have enough trust in the outcome that the model predicted for a particular claimant. Besides, they wanted to understand why one claimant has a higher probability of experiencing a future improvement in work ability than another claimant.

In general, a prediction based on fairly simple calculations is more likely to be used in practice than a model based on complex statistics [43]. However, there is a trade-off between interpretability and accuracy when choosing a type of prediction model. This balance is graphically illustrated in figure 2. This figure shows that regression models are among the solutions that are best interpretable, but also generally have the lowest accuracy in making predictions. By holding on to regression models, one commonly sacrifices accuracy for interpretability. In contrast, more advanced machine learning models with non-linear and non-smooth relations can make better predictions, but are also more difficult to interpret. Such models are considered black boxes as they provide only little insight into how the predictions were obtained from the input variables. For professionals, the interactions between the independent variables of such black box models can be too complex to understand and do often not make practical sense. When considering statistical models that are developed to be actually used in practice, one should keep the interpretability-accuracy trade-off in mind. In other words, given two models, one interpretable but less accurate, and the other non-interpretable but highly accurate, which one do they prefer?

The ability to make accurate and reliable predictions is often the most important feature of a prediction model. Full understanding of a model might be preferred, but is not a requirement from a data analytics perspective. To rely on the decisions made by prediction models, users generally need to have trust in the models' predicted outcomes. In this context, the concept of interpretability seems to be closely related to trust [44, 45]. Hence, if we could provide another source of trust in statistical models, the limitations of low interpretability could be overcome. An alternative way of building trust in prediction models is by comprehensive testing by IPs.

In this regard, several considerations should be taken into account. As prediction models are developed to be applied to new individuals/patients/claimants, their practical value depends on their performance outside the development sample, i.e. the sample that was used to build the model. Analysis of a model in the derivation sample results in performance indicators that are severely optimistic and more reliable performance indications can be obtained by using validation methods [46]. As shown in figure 3, bootstrapping or split-sample analysis are internal validation methods that correct for over-optimism and provide more accurate estimates of model performance [15, 47]. However, as (part of) the same dataset is used for model development and internal validation, internal validation does not give information about the model's performance in new cases. To this end, external validation of the model in a different sample of sufficient size is needed [48]. External validation gives an impression of whether model predictions are correct in different settings, such as for more recent cases or for different regions [49, 50]. Currently, the majority of the clinical prediction models never undergoes an external validation, which hinders conclusions about their stability and generalizability [51]. For the prediction models that we developed in chapters 3 and 4, we used internal validation methods to provide estimates that are more accurate. However, for practical and ethical reasons, we were unable to externally validate our models, and could therefore not make any statements about their use in practice. More frequent and thorough validation could be an effective way to build trust in prediction models and increase their impact in practical decision-making.

Moreover, although this is currently not the case, part of the design of our experimental study was that IPs would get their own caseload, i.e., if possible IPs would perform re-assessments for the claimants for which they also performed the initial work disability assessments. In this way,

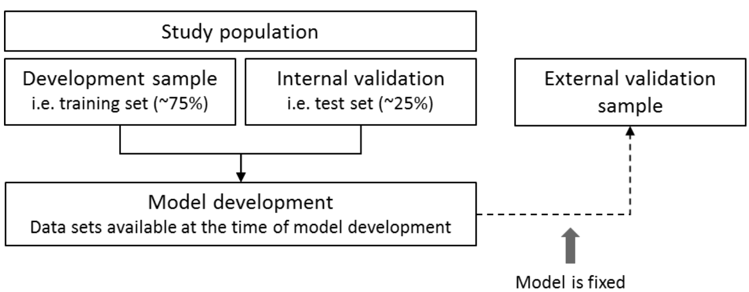


Figure 3. Data involved in internal and external validation

IPs would gain insight into the extent to which their own prediction of prognosis and the prognosis predicted by the model came true. This could enhance professionals' expertise about prognosis and could be a means of building trust in the prediction model.

Insurance physicians versus statistical prediction?

Often, physicians consider their own predictions and prognoses to be more accurate than those of prediction models. However, meta-analyses comparing clinical predictions with statistical predictions have shown the opposite [52, 53]. On average, statistical predictions were 10-13% more accurate than predictions made by clinicians. This superiority holds true for different types of mental and physical criteria, different settings, and was independent of the type of clinician and clinicians' working experience. Although the overall effect was moderate, it was consistent. In almost half of the studies, statistical predictions substantially outperformed clinicians' predictions, and in most of the other studies their accuracy was about the same. Only very few studies favored clinicians' predictions. These results imply that clinicians' time might best be used when spent on individual support of patients, whereas data analytics could be involved in prediction of diagnosis and prognosis [54]. Although there are no similar studies comparing professionals' predictions with statistical predictions in insurance medicine, similar results could potentially also hold for prediction of prognosis of work ability and classification of work disability claimants. As IPs generally consider more/other factors than the ones that are included in a prediction model, statistical predictions will be of most benefit when they are combined with IPs own predictions of prognosis. For instance, although on average statistical predictions might be more accurate for common cases, IPs can be expected to be better at predicting prognosis for claimants with specific and non-regular disorders and functional limitations.

To evaluate the benefits of prediction models for insurance medicine, the outcomes of these models should be compared with an alternative that is considered to be current best practice in the field of sickness absence and work disability assessments. Studying prediction models in occupational health practice, this would imply comparing the predicted outcomes with IPs' predictions on prognosis of work ability and return to work. This could be done using a randomized controlled trial (RCT) that relates both the IPs' own prognosis and the prognosis estimated by the prediction model to claimants' actual change in work ability at one-year follow-up. Combining an RCT with a process evaluation would give insight in the implementation and the views of IPs, and would help to interpret the outcome results. As mentioned earlier, we were unfortunately unable to conduct a process evaluation and an RCT evaluating the use of the prediction model and the decision support tool in practice.

However, performing an experimental study, preferably using a pragmatic RCT as study design, is the only way to get insight in the benefit of the decision support tool in daily practice. As described above, prediction models are in most cases more accurate than or at least as accurate as predictions made by professionals. Therefore, would it not be beneficial to combine these two sources of information? If IPs take part in experimental studies performed in everyday insurance medicine practice, they can get familiar with the use of decision support tools, and experience how it can help them in making more evidence-based decisions. From a claimants' perspective, using prediction models to assist IPs in making more accurate prognosis might be beneficial and this could also result in more effective allocation of return to work interventions. Moreover, as an invitation for a re-assessment generally results in feelings of uncertainty, preventing re-assessments that do not result in a change in the work disability benefit or return to work support might be in favor of the wellbeing of claimants. Hence, it is recommended to take claimants' perspectives into account in the experimental study as well.

Prediction models' fairness

An argument that was often brought up by SSI professionals during discussions about future use of a prediction model and decision support tool was that they felt that statistical models are discriminating. Professionals considered the human decision-making process to be more fair and more neutral than that of statistical models. They felt that human decision-making was transparent, while that of models was inscrutable and arbitrary. However, it could be argued that the opposite might be true.

Our minds are a hidden mental world of judgments, feelings and motives that steer our behavior [55]. Most choices that we make happen automatically, and we are generally unaware of why and how these choices have been made [56]. Research from behavioral science has demonstrated that even the most well-intentioned people may carry unconscious or hidden biases against certain groups from a lifetime of exposure to cultural attitudes about age, gender, ethnicity, social class, disability status, and so on [57]. Prediction models could be a guide to gain more insight in the rather inscrutable process of human decision-making. In prediction models, we know which data was used to develop the model, which screening rule is used, and how the outcome of the screening's rule would change if a specified prognostic factor would be changed. Hence, it could be argued that the well-regulated process of statistical models might be more transparent and less discriminating than the human decision-making process. To illustrate, statistical predictions are not influenced by external factors. Given the same input variables, these models will produce the same outcome, no matter what. Moreover, it could even explicitly be defined if certain variables should be omitted or included as prognostic factors. In our models, we used statistical methods for variable selection. The results of chapter 3 show that the division of claimants in the four subgroups was only based on health-related and psychosocial factors. The sociodemographic variables that were considered in the analyses (age, gender, educational level and marital status) were not selected by the statistical methods to be included in the final model. Hence, it would not be possible to favour or harm certain groups of claimants with respect to the choice of rehabilitation strategies based on characteristics such as their gender or socioeconomic status.

However, some caution is needed as prediction models are based on data of decisions that have been made in the past, and these data can be biased [58, 59]. Consider a model that can be used at the HR department of a company to select applicants for a job interview. Imagine that this model shows favourable predictions for men compared to women. In this case, we should not blame the model to harm women. Instead, as historical data of HR professionals' decisions on job applicants was used to train the model, this would show that during the past years HR professionals have been prejudiced against women. A similar reasoning holds for decisions in insurance medicine, for instance when data analytics give insights in which claimants will benefit most and should be offered a certain rehabilitation program.

When such discriminatory patterns become clear, prediction models could even be used to increase fairness [60]. At first, one may think that this could be achieved by excluding certain variables from the prediction model. However, a solution to control for the discriminative effect of biased data could be to take variables such as gender and ethnicity explicitly into account in the model [61, 62]. That way, the model can recognize that individuals belonging to a certain group get less favorable model outcomes than others, and can adjust for this disparity. For instance, regarding the example of the previous paragraph, as an alternative of selecting the top 10 interviewees from the total set of job applicants, we can agree to select the top 5 among the set of male applicants and the top 5 among the set of female applicants. Given that we aim for a fair solution in terms of gender, the predicted outcome would still represent the best possible ranking of potential applicants. In this way, including variables such as gender and ethnicity in data analytics can potentially improve fairness and reduce discrimination in the decision-making process. This shows that decisions made by prediction models could be considered more transparent and less discriminating than the human decision-making process.

Introducing prediction models in daily practice

To enhance successful implementation of prediction models and decision support tools in practice, some decisions need to be made depending on the future users of the tool and its application. This section describes different types of data analytics, different types of decision support tools, and their impact on professionals' decision-making.

Types of data analytics

There are different types of data analytics, varying in the complexity of the analysis and their benefit. Figure 4 shows the four types of data analytics that theoretically can be used to improve decision-making [63]. The vertical axis shows the added-value contribution, the horizontal axis shows the complexity that is needed to attain this value. The more complex an analysis is, the more value it brings, at least in theory. At the lowest levels, descriptive analytics answer the question of what happened, and diagnostic analytics try to interpret and explain why something happened. These analytics were used in chapter 2, in which we used registration data of the SSI to describe characteristics of work disability entitlement and continuing eligibility. Going one step further, predictive analytics describe what could happen and prescriptive analytics describe what should happen. We used predictive analytics in chapters 3 and 4 to estimate risk of long-term sickness absence and future changes in work ability.

These four types of data analytics for decision-making are usually implemented in practice in stages. They are complementary, and in some cases additive, i.e., the more sophisticated analytics cannot be applied without using the more fundamental analytics first. Descriptive and diagnostic analytics give valuable insights into the past, and are relatively easy to apply and explain. Subsequently, predictive analytics allow social security institutes to better understand patterns of sickness absence and work disability, and use that knowledge to proactively respond future trends. Hence, descriptive and diagnostic analytics are a good starting point for data analyses as they provide valuable insights. However, it is beneficial to take it one step further as in general predictive analytics leverage process optimization and effective allocation of scarce resources. Doing so, it is important to extensively test and validate these models before actually

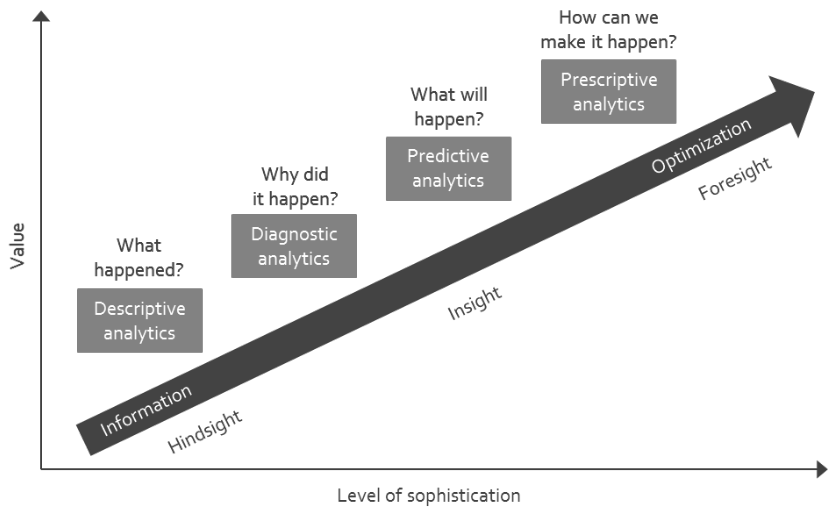


Figure 4. Four types of data analytics to improve decision-making

applying them in practice, to make sure they show good performance and they meet practical needs.

Impact of decision support tools

The prediction models and decision support tool described in this thesis are not meant as stand-alone tools, but aim to support professionals' decision-making. Professionals can discretionary act upon the predicted outcome of the tool, but this is far from mandatory. From chapter 5, we concluded that most professionals at the SSI prefer to first make their own prognosis and afterwards verify or adjust their evaluation based on the outcome of the decision support tool. However, the results of the efficacy study showed that, despite the fact that in more than half of the cases (53%) the prognosis of the IP was not in line with the outcome of the decision support tool, in only a minority of these cases (22%) IPs changed their prognosis after evaluating the tool (chapter 6). Given this finding, presenting the tool at the beginning of the decision-making process could probably increase its impact and feasibility in practice. In that way, IPs could assign the same value to the outcome of the tool as they do to other sources of information (e.g. guidelines, protocols and scientific literature). They could take the outcome of the tool into account when constructing a hypothesis about a claimant's work ability at the beginning of the work disability assessment process, and test this hypothesis during later stages of the process, e.g. during the interview with the claimant. This could potentially increase the efficacy of the tool.

Whether a decision support tool is implemented as an assistive tool (providing predicted outcomes without recommending decisions) or an advisory tool (explicitly recommending decisions) also determines the potential impact of the tool [64]. The data analyses solutions presented in chapters 3 and 4 belong to the former category: they present the predicted risk of long-term sickness absence or the predicted change in work ability at one-year follow-up, but no specific actions or recommendations are linked to these predictions. They could be transformed into an advisory tool by deducing specific recommendations on return to work interventions or scheduling of re-assessments from the predicted categories of sick-listed workers or the predicted changes in work ability. IPs may consider assistive decision support tools to be less threatening to their autonomy and more respectful of their professional judgement than advisory tools. However, previous research indicated that advisory decision support tools are more likely to result in behavioral change and are generally more effective [65, 66]. A study comparing an assistive and advisory format of the same prediction model within a single setting, demonstrated that an advisory format not only had a greater impact on clinical practice but also on patient outcomes [67]. When first implemented, organizations can decide to start with assistive tools to allow professionals some time to get used to working with decision support tools. In a next stage, it is recommended to gradually convert these to more advisory tools to fully benefit from the potential value of these tools in the decision-making process.

Another aspect concerns whether professionals should be obliged to consult a decision support tool in all cases, or whether this is left to their professional judgement. The former could improve the transparency and uniformity of IPs' decisions regarding prognoses, but could also result in higher resistance against use of the tool. As presented in chapter 5, the participants of our focus groups agreed that they would not beforehand exclude certain types of claimants from the tool, as there is always some uncertainty around the prognosis. They argued that the tool is always a useful complement to their own prognosis. Depending on how unsure they are about the prognosis, they would use it either to establish prognosis or verify their thoughts. However, in the efficacy study presented in chapter 6, some IPs mentioned that they are reluctant consulting the tool for claimants for which prognosis is evident to them. If use of the tool would be mandatory, still including these claimants would undermine IPs' willingness to use the tool. Nevertheless, making consultation of the tool compulsory is expected to improve successful implementation in practice as it reinforces actual integration of the tool in the standard working and decision-making process. In this regard, automatic provision of the outcome of the prediction model and smooth integration within the current workflow of IPs are key factors [68]. The decision support tool is certainly not meant as a stand-alone tool, but as a tool to support IPs' decision-making. Hence, professionals, even if they would be obliged to consult the tool, would not be obliged to follow its outcome. IPs can always deviate from the outcome of the prediction model. Insight in cases for which IPs' prognosis deviates from the outcome of the prediction model and why this is the case could be used to improve the tool to better suit professionals' expertise.

Recommendations for research, policy and practice

Considering the themes that were described and discussed in this chapter, several recommendations for research and practice can be derived from the findings of this thesis.

Recommendations for research

- Prediction models for long-term sickness absence and work disability can be useful to enhance evidence-based decision-making. When developing such models, it is recommended to combine registration data of social security institutes with data from self-reported questionnaires. It was shown that they both contain important prognostic factors and combining them can improve the accuracy of the prediction model. Moreover, to develop models that are actually useful in practice, researchers should have a close look at the data and the statistical techniques that are most suitable. For instance, if only a minority of the study population experiences the event of interest, standard logistic regression techniques might be unable to identify these individuals. Remedies for this problem include over- or under-sampling, weighted regression, applying penalization methods, or applying a skewed link function.

- Although numerous prediction models are constantly being developed in the literature, very few of them are actually used in practice. The majority of the models are only described once and never validated. To be useful in practice, a model should be credible, accurate and have shown its clinical effectiveness. Hence, external validation of a prediction model is essential to gain insight into the performance of the model in a sample of future cases and into its benefit for practice. Instead of using newly collected data to develop new prediction models, they should rather be used to validate, adjust, and update existing models.
- Once validated, impact studies should subsequently quantify the effect of using prediction models on professionals' decision-making, patient outcomes and cost-effectiveness. Impact studies require a control group of healthcare professionals who provide usual care. Because of legal, ethical and practical limitations, such studies are often difficult to execute. However, they are a necessary step in confirming that a well performing prediction model also improves professionals' decision-making.
- In addition to quantitatively assessing the impact of a prediction model on professionals' behavior, qualitative studies are a valuable tool to assess professionals' attitudes towards such models. Process evaluations can give insight into professionals' evaluations of the use of prediction models in practice and guide directions for improvement. Moreover, they can provide valuable information for the development of effective strategies for the implementation of decision support tools in practice.
- Research is now mostly focused on analyses of structured data, i.e. data that is clearly defined and easily analyzed. On the contrary, unstructured data contains information that is not easily accessible to computational data management. These include, for instance, large amounts of written text data embedded in medical reports, telephone calls or recordings of interviews with claimants, and/or web data. In the last decade, there has been a dramatic increase in the amount of unstructured data and these data can offer a wealth of information relevant to understanding sickness absence and work disability. Text mining and natural language processing should be explored as a way to organize large sets of unstructured data and extract meaningful information from them.

Recommendations for policy

- Social security institutes can use their data to gain additional insights to improve decision-making, working processes, and services to claimants. To do so, it is important that professionals working at these institutes are aware of data quality issues. Systematically defining, characterizing and improving data quality is key to valid, generalizable and precise data analytics insights and predictions.

- Occupational health professionals and policy makers should take notice of the factors found in this thesis to be associated with long-term sickness absence and work disability. These characteristics could be addressed during the work disability assessment interview with a claimant, and could be used to offer tailored return to work interventions.
- Not only health-related factors but also other variables, such as self-reported psychological factors and expectations about return to work, were found to be important prognostic factors for long-term sickness absence and work disability. Nowadays, self-reported data can relatively easily be gathered with online questionnaires. Therefore, it is recommended to use online standardized questionnaires that all claimants who apply for a sickness absence or work disability benefit need to fill in. Short questionnaires are preferred in terms of costs and missing data. The results of this thesis give an indication of the questions that should be addressed.
- For successful future implementation of decision support tools within the field of insurance medicine, the SSI should be clear about the advantages and challenges of data analysis, and should anticipate barriers to implementation, so that strategies to minimize the impact of potential barriers or avoid them altogether can be developed. This will help to create an environment in which professionals are willing to learn about the possible advantages of decision support tools. Starting with a non-compulsory, assistive tool will most likely lead to the path of least resistance, and gives professionals the opportunity to explore the benefits of data analytics in an approachable way.
- When developing decision support tools, management and professionals should be involved from the beginning and continuously help to tailor the tool to the specific needs and context of future users.
- There is a limited capacity of occupational health resources, and especially a shortage in the number of IPs at the SSI. Prediction models and decision support tools can help to use the limited available capacity as effective as possible, i.e. to plan re-assessments for claimants who will benefit most from contact with an IP and for effective allocation of return to work support. This may not only be beneficial from an organizational perspective, but also for individual professionals and claimants, as professionals can spend their time on activities and on claimants for whom this is of most benefit.

Recommendations for practice

- IPs predominantly rely on sociodemographic and health-related factors, such as age and diagnoses, when conducting sickness absence and work disability assessments [69, 70]. However, as we have shown in chapters 3 and 4, more subjective measures, such as self-reported expectations about return to work, coping strategies, health experience, and social support from relatives, can be important factors for sickness absence and work disability duration as well. It is recommended to take this broad range of factors into account during sickness absence and work disability assessments.
- Professionals are often worried about the use of decision support tool in practice. They might fear that such tools will take over their jobs, and claimants might fear that they will lead to automatically generated decisions regarding their benefit and return to work support. Clear communication is important to show that these fears are unjustified. For instance, it should be emphasized that prediction models are certainly not meant to take over the job of professionals but to support them to make more evidence-based decisions. In addition, reasonable explanation about the data, the analysis and the application will increase willingness to explore the use of decision support tools in practice. This could be done by conducting information sessions on these topics, or by incorporating them in professionals' training programs.
- To get insights in the potential benefit of decision support tools, professionals should be willing to experiment using these tools in practice, for instance in an RCT. This enables answering questions that otherwise cannot be answered. For instance, were the estimated changes in work ability of the prediction model in agreement with IPs' own prognosis? When evaluated at one-year follow-up, how accurate were IPs' own prognoses and the statistical predictions? Can we distinguish specific types of claimants for whom the prediction model might in particular be helpful? Hence, such small scale experimental studies would give valuable insights. As decision support tools are meant as auxiliary tools to assist professionals' decisions-making, and certainly not meant to take over their jobs, participation in such experiments will be of no harm to professionals' autonomy. Hence, we recommend professionals to sincerely consider participation in experimental studies.

Conclusions

Prediction models and decisions support tools provide early identification of individuals at risk of long-term sickness absence and work disability, and give insight in prognostic factors. This way, data analytics can help IPs and other occupational health professionals in making evidence-based decisions regarding prognosis of functional abilities and support to return to work. Prediction models should be externally validated and evaluated with impact studies and process evaluations to assess, in practice, their feasibility and performance on professionals' decision-making. Clear communication about the development and content of the prediction model, and about the benefits and challenges of data analytics in general, would facilitate successful implementation of such tools in practice. As IPs that took part in the efficacy study mentioned:

"Before I decide to use this tool in daily practice, I want to know how well it performs when it is validated with data of new claimants."

"I hope that tools like these will also become available for other professionals and can be applied to other types of claimants."

"Digital support tools and machine learning algorithms can help us in identifying high-risk groups and allocation of scarce resources. I hope that they will assist me in the future in making more evidence-based decision."

References

1. Weerdesteijn KH, Schaafsma FG, Louwerse I, Huysmans MA, Van der Beek AJ, Anema JR. Does self-perceived health correlate with physician-assessed functional limitations in medical work disability assessments? J Psychosom Res. 2019;125:109792.
2. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc. 2016;9:211-217.
3. Rosenman R, Tennekoon V, Hill LG. Measuring bias in self-reported data. Int J Behav Healthc Res. 2011;2(4):320-332.
4. Dowrick AS, Gabbe BJ, Williamson OD, Wolfe R, Cameron PA. A comparison of self-reported and independently observed disability in an orthopedic trauma population. J Trauma Acute Care Surg. 2006;61(6):1447-1452.
5. Marfeo EE, Eisen S, Ni P, Rasch EK, Rogers ES, Jette A. Do claimants over-report behavioral health dysfunction when filing for work disability benefits? Work. 2015;51(2):187-194.
6. Brouwer S, Dijkstra PU, Stewart R, Göeken L, Groothoff J, Geertzen J. Comparing self-report, clinical examination and functional testing in the assessment of work-related limitations in patients with chronic low back pain. Disabil Rehabil. 2005;27(17):999-1005.
7. Dell-Kuster S, Lauper S, Koehler J, Zwimpfer J, Altermatt B, Zwimpfer T, et al. Assessing work ability-a cross-sectional study of interrater agreement between disability claimants, treating physicians, and medical experts. Scand J Work Environ Health. 2014;40(5):493-501.
8. Slebus F, Kuijer P, Willems J, Frings-Dresen M, Sluiter J. Work ability assessment in prolonged depressive illness. Occup Med. 2010;60(4):307-309.
9. Wind H, Samoocha D, van der Beek AJ, Frings-Dresen MH. Prevention of disability: The opinion of claimants applying for a disability benefit. Work. 2014;49(2):335-341.
10. Nieuwenhuijsen K, Cornelius L, de Boer M, Groothoff J, Frings-Dresen M, van der Klink J, et al. Predicting improvement of functioning in disability claimants. J Occup Rehabil. 2014;24(3):410-418.
11. Howard GS. Response-shift bias: A problem in evaluating interventions with pre/post self-reports. Eval Rev. 1980;4(1):93-106.
12. Van Egmond M, Duijts S, Loyen A, Vermeulen S, Van der Beek A, Anema J. Barriers and facilitators for return to work in cancer survivors with job loss experience: a focus group study. Eur J Cancer Care. 2017;26(5):e12420.
13. Audhoe SS, Nieuwenhuijsen K, Hoving JL, Sluiter JK, Frings-Dresen MH. Perspectives of unemployed workers with mental health problems: barriers to and solutions for return to work. Disabil Rehabil. 2018;40(1):28-34.
14. Mullainathan S, Spiess J. Machine learning: an applied econometric approach. J Econ Perspect. 2017;31(2):87-106.
15. Harrell FE. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Switzerland: Springer International Publishing; 2015.
16. Martin-Sanchez F, Verspoor K. Big data in medicine is driving big changes. Yearb Med Inform. 2014;23(1):14-20.
17. OECD. Transforming disability into ability: Policies to promote work and income security for disabled people. Paris: OECD Publishing; 2003.
18. Anema JR, Schellart AJ, Cassidy J, Loisel P, Veerman T, Van der Beek A. Can cross country differences in return-to-work after chronic occupational back pain be explained? An exploratory analysis on disability policies in a six country cohort study. J Occup Rehabil. 2009;19(4):419-426.

19. Lippel K, Lötters F. Public insurance systems: a comparison of cause-based and disability-based income support systems, in Handbook of work disability. 2013, Springer: New York.
20. Hollon TC, Pandian B, Adapa AR, Urias E, Save AV, Khalsa SSS, et al. Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. Nat Med. 2020;26(1):52-58.
21. Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. Lancet Oncol. 2020;21(2):233-241.
22. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature. 2020;577(7788):89-94.
23. Blanks R, Wallis M, Given-Wilson R. Observer variability in cancer detection during routine repeat (incident) mammographic screening in a study of two versus one view mammography. J Med Screen. 1999;6(3):152-158.
24. Gilbert FJ, Astley SM, Gillan MG, Agbaje OF, Wallis MG, James J, et al. Single reading with computer-aided detection for screening mammography. N Engl J Med. 2008;359(16):1675-1684.
25. De Dombal F, Leaper D, Horrocks JC, Staniland JR, McCann A. Human and computer-aided diagnosis of abdominal pain: further report with emphasis on performance of clinicians. Br Med J. 1974; 1(5904):376-380.
26. Gambrill E. Critical thinking in clinical practice: Improving the quality of judgments and decisions. New Jersey: John Wiley & Sons; 2006.
27. Kok R, Hoving JL, Verbeek J, Schaafsma FG, van Dijk FJ. Integrating evidence in disability evaluation by social insurance physicians. Scand J Work Environ Health. 2011;37(6):494-501.
28. de Boer W-E, Besseling J, Willems J. Organisation of disability evaluation in 15 countries. Prat Organ Soins. 2007(3):205-217.
29. de Jong P, Everhardt T, Schrijvershof C. Duurzaam niet-duurzaam? Onderzoek naar niet-duurzaam volledig arbeidsongeschikt verklaarden, in Ape rapport nr. 967. 2013: Den Haag.
30. Rebergen DS, Bruinvels DJ, Bos CM, van der Beek AJ, van Mechelen W. Return to work and occupational physicians' management of common mental health problems-process evaluation of a randomized controlled trial. Scand J Work Environ Health. 2010;36(6):488-498.
29. van der Weide WE, Verbeek J, Van Dijk F. Relation between indicators for quality of occupational rehabilitation of employees with low back pain. J Occup Environ Med. 1999;56(7):488-493.
30. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ. 2009;338:b375.
31. Waddell G, Burton AK, Kendall NA. Vocational rehabilitation–what works, for whom, and when? Report for the Vocational Rehabilitation Task Group. London: The Stationery Office; 2008.
32. Barth J, de Boer WE, Busse JW, Hoving JL, Kedzia S, Couban R, et al. Inter-rater agreement in evaluation of disability: systematic review of reproducibility studies. BMJ. 2017;356:j14.
33. Kobak KA, Brown B, Sharp I, Levy-Mack H, Wells K, Ockun F, et al. Sources of unreliability in depression ratings. J Clin Psychopharmacol. 2009;29(1):82-85.
34. Schellart AJ, Mulders H, Steenbeek R, Anema JR, Kroneman H, Besseling J. Inter-doctor variations in the assessment of functional incapacities by insurance physicians. BMC Public Health. 2011; 11(1):864.
35. Kox RJ, Hoving JL, Verbeek JH, Schouten MJ, Hulshof CT, Wind H, et al. Assessment of prognosis by physicians involved in work disability evaluation: A qualitative study. PloS One. 2019;14(2).

36. Spanjer J, Krol B, Brouwer S, Groothoff JW. Inter-rater reliability in disability assessment based on a semi-structured interview report. Disabil Rehabil. 2008;30(24):1885-1890.
37. Eisenberg JM. Physician utilization the state of research about physicians' practice patterns. Med care. 2002;40(11):1016-1035.
38. Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury. BMC Med Inform Decis Mak. 2006;6(1):38.
39. Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. BMJ. 2011;343:d7163.
40. Hallen SA, Hootsmans NA, Blaisdell L, Gutheil CM, Han PK. Physicians' perceptions of the value of prognostic models: the benefits and risks of prognostic confidence. Health Expect. 2015;18(6):2266-2277.
41. Wyatt JC, Altman DG. Commentary: Prognostic models: clinically useful or quickly forgotten? BMJ. 1995;311(7019):1539-1541.
42. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608. 2017.
43. Herman B. The promise and peril of human evaluation for model interpretability. arXiv preprint arXiv:1711.07414. 2017.
44. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal–external, and external validation. J Clin Epidemiol. 2016;69(1):245-247.
45. Steyerberg EW. Clinical prediction models. Switzerland: Springer Nature; 2019.
46. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Annals of internal medicine. 1999;130(6):515-524.
47. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ. 2009;338:b605.
48. Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol. 2015;68(3):279-289.
49. Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol. 2015;68(1):25-34.
50. Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: a meta-analysis. Psychol Assess. 2000;12(1):19-30.
51. Aegisdottir S, White MJ, Spengler PM, Maugherman AS, Anderson LA, Cook RS, et al. The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. J Couns Psychol. 2006;34(3):341-382.
52. Meehl PE. Clinical versus statistical prediction: A theoretical analysis and a review of the evidence. Washington: American Psychological Associatino; 1954.
53. Wilson TD. Strangers to ourselves. Boston: Harvard University Press; 2004.
54. Nisbett RE, Wilson TD. Telling more than we can know: verbal reports on mental processes. Psychol Rev. 1977;84(3):231-259.
55. Banaji M, Greenwald A. Blindspot: Hidden Biases of Good People. New York: Delacorte Press; 2013.
56. Barocas S, Selbst AD. Big data's disparate impact. Calif L Rev. 2016;104:671-732.
57. Mayson SG. Bias in, bias out. Yale Law J. 2018;128(8):2218-2300.
58. Kleinberg J, Ludwig J, Mullainathan S, Sunstein CR. Discrimination in the Age of Algorithms. J Leg Anal. 2018;10:113-174.

59. Kleinberg J, Ludwig J, Mullainathan S, Rambachan A. Algorithmic fairness. in AEA Papers and Proceedings. 2018.
60. Gillis TB, Spiess JL. Big data and discrimination. Univ Chic Law Rev. 2019;86(2):459-488.
61. Davenport T, Harris J. Competing on analytics: The new science of winning. Boston: Harvard Business Press; 2017.
62. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Ann Intern Med. 2006;144(3):201-209.
63. Grimshaw JM, Shirran L, Thomas R, Mowatt G, Fraser C, Bero L, et al. Changing provider behavior: an overview of systematic reviews of interventions. Med care. 2001;39(8):2-45.
64. Michie S, Johnston M. Changing clinical behaviour by making guidelines specific. BMJ. 2004; 328(7435):343-345.
65. Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KG. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. Diagn Progn Res. 2018;2(1):11.
66. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ. 2005;330(7494):765.
67. Krause N, Frank JW, Dasinger LK, Sullivan TJ, Sinclair SJ. Determinants of duration of disability and return-to-work after work-related injury and illness: Challenges for future research. Am J Ind Med. 2001;40(4):464-484.
68. Slebus FG, Sluiter JK, Kuijer PPF, Willems JH, Frings-Dresen MH. Work-ability evaluation: a piece of cake or a hard nut to crack? Disabil Rehabil. 2007;29(16):1295-1300.

Appendix

Summary

There is a positive association between work and one's wellbeing, mental and physical health. Work disability is generally bad for an individual's health, and returning to work is generally associated with a positive effect on the future course of the disease and work ability. Moreover, long-term sickness absence and work disability are among the greatest social and labor market challenges for policy makers in most OECD countries, as these countries spend on average 2% of their GDP on these benefits. Hence, prevention of work disability and support for returning to work are in the interest of individuals and society as a whole.

In the Netherlands, individuals who are unable to work due to a disease or disorder can apply for a sickness absence or work disability benefit. This covers both financial support to compensate loss of income and interventions supporting return to work. Based on insurance physicians' (IPs) assessment of diagnoses and functional limitations, it is determined whether a benefit should be granted or not. During these assessments, prognosis of future changes in work ability is an important task of IPs as, once a benefit has been granted, changes in health may alter its continuing eligibility. However, IPs consider this task as one of the most difficult parts of the work disability assessment as it requires rather complex predictions, in which a broad range of factors play a role.

Data analytics and prediction models give an overview of factors associated with sickness absence and work disability duration, and can be used to target specific at-risk groups. They can help occupational health professionals to identify effective return to work interventions and ensure that medical re-assessments are planned at the time an assessment interview with an IP has most added value.

The main aim of this thesis was to investigate how data analysis, prediction models and decision support tools can help IPs in the Netherlands in making evidence-based decisions regarding planning of re-assessments and support to return to work. Specifically, the following research objectives were addressed in this thesis:

1. To give an overview of factors associated with work disability entitlement and duration;
2. To predict risk of long-term sickness absence and identify distinct subgroups of sick-listed workers without an employment contract;
3. To develop a prediction model and decision support tool predicting future changes in work ability of work disability claimants;
4. To get insight into the efficacy of the decision support tool and IPs' attitudes towards use of the tool.

Factors associated with long-term sickness absence and work disability

In **chapter 2**, the main diagnoses of workers who qualify for work disability benefits were examined, and it was explored how these diagnoses differed among age groups, gender and educational level. A cohort study of 31,733 individuals receiving work disability benefits from the Dutch Social Security Institute (SSI) showed that mental disorders were the most frequent diagnosis for individuals claiming long-term work disability. Diagnoses differed among age groups and education categories; whereas mental disorders were the main diagnosis for work disability among younger and more highly educated individuals, physical disorders were the main diagnosis among older and less educated individuals. Using a five-year follow-up, it was shown that claim duration for disability benefits was long lasting for most claimants. Continuing eligibility for disability benefit was highest for individuals with (multiple) mental disorders and those with a comorbidity of mental and physical disorders, and lowest for individuals with (multiple) physical disorders.

Chapter 3 aimed to predict risk of long-term sickness absence and identify distinct subgroups among sick-listed workers without an employment contract. A cohort of 437 individuals who were granted a sickness absence benefit for at least two weeks was followed for 1 year. For these individuals, registration data of the SSI was combined with self-reported questionnaires on sociodemographic, work-related, health-related and psychosocial factors. Based on educational level, self-reported expected sickness absence and help-seeking ability as prognostic factors, it was possible to fairly discriminate between individuals with and without long-term sickness absence. Subsequently, the predicted risk of long-term sickness absence was used in combination with self-reported variables to identify four subgroups: sick listed workers with mental limitations, sick listed workers with physical limitations, sick listed workers with positive expectations, and sick listed workers with negative expectations about their return to work. These findings could be used to identify individuals at risk of long-term sickness absence. In this way, they can aid professionals to offer tailored return to work interventions to the groups that will most benefit from it.

Development of a prediction model and decision support tool for changes in work ability

In **chapter 4**, weighted regression was used to predict changes in work ability one year after approval of the work disability benefit. The study population consisted of 944 individuals who were granted a work disability benefit by the SSI. For these individuals, self-reported questionnaire data measured at baseline were linked with administrative data from SSI databases. The results showed that there are indications that weighted regression procedures can correctly identify more individuals who experience a relevant change in WAS compared to standard multinomial logit models. The prediction model can assist IPs in identifying claimants with a high probability of experiencing an improvement of work ability at one-year follow-up.

This model can aid accurate prognosis of work ability, planning of re-assessments and provision of suitable interventions for return to work.

The prediction model resulted in a statistical formula that uses prognostic factors to predict changes in work-ability at one-year follow-up. Next, we designed a decision support tool, i.e. a suitable graphical user interface, which can be provided to IPs in order to use the prediction model in practice. The aim of **chapter 5** was to explore professionals' preferences regarding the way of use and design of this tool. A focus group study among IPs and labor experts of the SSI showed that clarity and ease of use were important features of the tool. Dividing claimants into categories based on the outcome of the prediction model and assigning color labels to the categories was experienced as the most straightforward and clear way of presenting the results of the prediction model. Concerning preferences on when to use the tool, most professionals stated that they would prefer to first make their own judgement during the work disability assessment interview with the claimant and afterwards verify or adjust their evaluation based on the outcome of the decision support tool. These features should be taken into account when developing the tool, in order to encourage professionals to use the tool in practice and act accordingly.

Added value of the decision support tool for insurance physicians

Chapter 6 focused on evaluating the efficacy of the decision support tool on IPs prognosis of work ability and their prognostic confidence. A vignette study among 29 IPs showed that the congruence of the decision support tool with IPs' prognosis of future work ability was low, and that IPs' prognostic confidence decreased after evaluating the tool if their prognosis was discordant with the outcome of the prediction model. Moreover, this study investigated professionals' attitudes towards use of the decision support tool in practice. IPs stated that the wish to know more about the tool was the main barrier for use. Although the perceived barriers were overall limited, only a minority of the IPs indicated that they would be willing to use the tool in practice or expected that their colleagues would be willing to do so. The findings of this study indicate that making professionals more familiar with prediction models and decision support tools is an important factor for successful future implementation.

Conclusion

This thesis showed that prediction models and decisions support tools provide early identification of individuals at risk of long-term sickness absence and work disability. These models and tools can help IPs and other occupational health professionals in making evidence-based decisions regarding prognosis of functional abilities, for planning of re-assessments and identification of effective return to work interventions. External validation of prediction models and decisions support tools is necessary to evaluate their added value in practice. To do so, researchers are dependent on the willingness of IPs and other professionals to participate in small scale experiments using these tools in practice. As decision support tools are meant as auxiliary tools to assist professionals' decisions-making, and certainly not meant to take over their jobs, participation in such experiments will be of no harm to professionals' autonomy. Conducting information sessions on data analytics and prediction models, or by incorporating them in professionals' training programs, could be an important step to make professionals more familiar with these topics.

Samenvatting

Mensen die werken hebben vaak een betere gezondheid, zowel fysiek als psychisch, dan mensen zonder betaald werk. Wanneer iemand die ziek is (geweest) weer aan het werk gaat, heeft dat vaak een positief effect op het ziekteverloop en toekomstig werkvermogen. Bovendien leiden ziekteverzuim en langdurige arbeidsongeschiktheid tot financiële nadelen voor de zieke werknemers, werkgevers en de maatschappij. Het voorkomen van langdurig ziekteverzuim en arbeidsongeschiktheid en het bieden van ondersteuning bij terugkeer naar werk zijn dan ook in het belang van zowel degene die ziek of arbeidsongeschikt is als de samenleving in zijn geheel.

Als mensen langdurig ziek zijn en daardoor hun werk niet meer (volledig) kunnen doen, kunnen zij een aanvraag doen voor arbeidsongeschikteidsuitkering (Ziektewet- of WIA-uitkering) bij UWV. Iemand die een uitkering aanvraagt, krijgt een uitnodiging om op het spreekuur van de verzekeringsarts te komen. Tijdens zo'n arbeidsongeschiktheidsbeoordeling kijkt de arts naar iemands gezondheidssituatie, en naar welke beperkingen iemand heeft door zijn of haar ziekte of aandoening. De arts beoordeelt ook of iemands klachten en beperkingen tijdelijk of blijvend zijn. Als de verzekeringsarts vaststelt dat iemand nog mogelijkheden heeft om te werken, volgt een uitnodiging voor een gesprek met de arbeidsdeskundige. Met de informatie van de verzekeringsarts kijkt de arbeidsdeskundige wat voor werk iemand nog kan doen. Op basis hiervan wordt bepaald of iemand recht heeft op een uitkering.

Als de gezondheid van mensen die een arbeidsongeschiktheidsuitkering ontvangen verbetert of verslechtert, kan dit invloed hebben op hun mogelijkheden om te werken. Verzekeringsartsen voeren daarom herbeoordelingen uit. Tijdens een herbeoordeling kijkt een arts opnieuw naar iemands gezondheidssituatie. Om een inschatting te kunnen maken of en op welke termijn een herbeoordeling zinvol is, is het belangrijk dat verzekeringsartsen een goede prognose stellen over of en wanneer er iets verandert aan iemands gezondheid. Verzekeringsartsen beschouwen de prognose echter als een van de lastigste aspecten van hun werk. Bij het stellen van een prognose gaat het namelijk vaak om complexe voorspellingen waarbij veel verschillende factoren een rol spelen.

Data analytics en voorspelmodellen kunnen inzicht geven in de factoren die samenhangen met de duur van de arbeidsongeschiktheidsuitkering. Om een voorspelmodel te kunnen gebruiken in de praktijk, moet het onderdeel worden van een beslissingsondersteunend instrument. Een belissingsondersteunend instrument is een (online) hulpmiddel dat de uitkomsten van een voorspelmodel samenvat en op een duidelijke manier presenteert aan de gebruiker. Op deze manier kunnen data analytics UWV-professionals, zoals verzekeringsartsen, helpen bij het stellen van een goede prognose. Dit draagt bij aan het gericht inzetten van dienstverlening voor terugkeer naar werk en het doelgericht plannen van herbeoordelingen.

Het hoofddoel van dit promotieonderzoek was om te onderzoeken hoe data analytics, voorspelmodellen en beslissingsondersteunende instrumenten verzekeringsartsen kunnen helpen bij het maken van (wetenschappelijk) onderbouwde beslissingen rondom het plannen van WIA-herbeoordelingen en de inzet van ondersteuning voor terugkeer naar werk. Dit hoofddoel is uitgesplitst naar vier subdoelen:

1. Een overzicht geven van factoren die samenhangen met het krijgen van een WIA-uitkering en de duur van de uitkering;
2. Voorspellen welke mensen een groter risico hebben om langdurig een Ziektewet-uitkering te ontvangen en het indelen van verzuimers in subgroepen;
3. Ontwikkelen van een voorspelmodel en beslissingsondersteunend instrument voor toekomstige veranderingen in werkvermogen van mensen die een WIA-uitkering aanvragen;
4. Inzicht krijgen in de toegevoegde waarde van het beslissingsondersteunend instrument en hoe verzekeringsartsen aankijken tegen gebruik van het instrument in de dagelijkse praktijk.

Factoren die samenhangen met langdurig ziekteverzuim en arbeidsongeschiktheid

In **hoofdstuk 2** hebben we gekeken wat de belangrijkste diagnoses zijn op basis waarvan mensen een WIA-uitkering krijgen. Uit de UWV-systemen hebben we data verzameld van 31.733 mensen (o.a. socio-demografische kenmerken, gezondheidssituatie en werkgerelateerde factoren). De diagnosecategorie die verzekeringsartsen het vaakst registreerden bij toekenning van een WIA-uitkering was psychische klachten. Diagnoses verschilden tussen mensen met verschillende leeftijden en opleidingsniveaus; psychische klachten werden vaker genoemd bij jongeren en hoger opgeleiden, lichamelijke klachten waren de belangrijkste diagnose voor ouderen en lager opgeleiden. Daarnaast hebben we een follow-up periode van vijf jaar gebruikt om te onderzoeken wat de kenmerken waren van mensen die na vijf jaar nog steeds een WIA-uitkering ontvingen. Hieruit bleek dat 82% van de mensen na 5 jaar nog steeds een uitkering ontving. Mensen met (meerdere) psychische klachten en mensen met een combinatie van psychische en lichamelijke klachten ontvingen relatief vaak nog steeds een uitkering aan het eind van de vijf-jaar follow-up periode.

Het doel van **hoofdstuk 3** was het voorspellen van langdurig ziekteverzuim en het indelen van mensen in de Ziektewet in subgroepen met gelijke kenmerken. Een groep van 437 mensen die een Ziektewetuitkering toegekend kregen en die langer dan twee weken ziek waren, is een jaar

A

gevolgd. Op basis van opleidingsniveau, eigen verwachting ten aanzien van werkhervatting en het zelfgerapporteerde vermogen om anderen om hulp te vragen, konden we voorspellen welke mensen een hoger risico hadden om langdurig een Ziektewetuitkering te ontvangen. Het voorspelde risico werd in combinatie met verschillende zelfgerapporteerde factoren gebruikt om deze mensen in te delen in vier subgroepen. Hieruit bleek dat mensen in de Ziektewet met klachten van het bewegingsapparaat, die het makkelijk vinden om anderen om hulp te vragen en die positieve verwachtingen hebben ten aanzien van werkhervatting, een relatief laag risico hebben op langdurig verzuim. Mensen met psychische klachten die bij hun ziekmelding het gevoel hebben weinig controle te hebben over de dingen die hen overkomen en die negatieve verwachtingen hebben, hebben juist een hoger risico op langdurig verzuim. Deze resultaten kunnen gebruikt worden om te kijken of het wenselijk is om voor de gevonden groepen gerichte dienstverlening in te zetten die terugkeer naar werk bevorderen.

Een voorspelmodel en instrument voor veranderingen in werkvermogen

In **hoofdstuk 4** is de analysemethode 'gewogen regressie' gebruikt voor het voorspellen van veranderingen in werkvermogen één jaar na toekenning van de WIA-uitkering. De onderzoekspopulatie bestond uit 944 mensen die vlak voor de aanvraag van de uitkering, en één jaar daarna, een uitgebreide vragenlijst hadden ingevuld. In de vragenlijst konden mensen bijvoorbeeld aangeven welke klachten zij ervaarden, op welke manier dit hen belemmerde in hun werk en hoe zij aankeken tegen terugkeer naar werk. De antwoorden die mensen gaven hebben we gekoppeld aan registratiedata van UWV. Ons voorspelmodel bepaalt vervolgens voor mensen die een WIA-uitkering aanvragen of de kans het grootst is dat hun werkvermogen het komende jaar zal verbeteren, gelijk zal blijven of zal verslechteren. Deze voorspelling kan verzekeringsartsen helpen bij het stellen van een goede prognose. Hiermee kan het model bijdragen aan het doelgericht plannen van herbeoordelingen en het inzetten van gerichte begeleiding voor terugkeer naar werk.

Om het statistische voorspelmodel te kunnen gebruiken in de praktijk moet het onderdeel worden van een beslissingsondersteunend instrument. Het doel van **hoofdstuk 5** was inzicht krijgen in hoe dit instrument eruit moet komen te zien en hoe professionals het zouden willen gebruiken. Uit focusgroepen met verzekeringsartsen en arbeidsdeskundigen van UWV bleek dat duidelijkheid en gebruiksgemak belangrijke eigenschappen zijn voor het instrument. De meest eenvoudige en heldere manier om de resultaten van het voorspelmodel te presenteren is door mensen op basis van de uitkomst van het voorspelmodel in te delen in groepen. Vervolgens kunnen aan deze groepen kleuren toegewezen worden; bijvoorbeeld groen als er een grote kans is op verbetering en rood als er een grote is kans is op verslechtering van het werkvermogen. Als het gaat om voorkeuren voor het moment waarop het instrument gebruikt zou moeten worden, hadden professionals veelal de voorkeur om eerst zelf een prognose te stellen. Na afloop van het beoordelingsgesprek met de klant zouden artsen dan hun eigen prognose willen vergelijken met de uitkomst van het voorspelmodel. De resultaten van dit kwalitatieve onderzoek zijn meegenomen bij de ontwikkeling van het instrument, zodat professionals aangemoedigd worden om het instrument ook daadwerkelijk te gebruiken en de uitkomst mee te wegen in hun oordeel.

Toegevoegde waarde van het instrument voor verzekeringsartsen

De resultaten van de evaluatie van het gebruik van het beslissingsondersteunend instrument door verzekeringsartsen staan in **hoofdstuk 6**. In de evaluatie is gekeken of het instrument helpt bij het stellen van een goede prognose en welke invloed het heeft op het vertrouwen dat verzekeringsartsen hebben in hun eigen prognose. Dit hebben we gedaan met een vignetten-studie, waarin 29 verzekeringsartsen zes casussen van mensen met een WIA-uitkering hebben beoordeeld. Deze studie liet zien dat de uitkomst van het instrument (verbetering/ver-slechtering/geen verandering in werkvermogen) in meer dan de helft van de gevallen (53%) niet overeenkwam met de eigen prognose van de verzekeringsarts. Als de prognose van de verzekeringsarts niet overeenkwam met de uitkomst van het instrument, daalde bovendien zijn of haar vertrouwen in de eigen prognose na het raadplegen van het instrument. Daarnaast werd in dit hoofdstuk gekeken hoe professionals aankijken tegen gebruik van het instrument in de dagelijkse praktijk. Hoewel er in het algemeen weinig belemmeringen werden ervaren voor gebruik, gaf slechts iets meer dan een kwart van de verzekeringsartsen aan het instrument te willen gebruiken in de dagelijkse praktijk. Ondanks een toelichting bij de start van het experiment over (de factoren in) het voorspelmodel, gaven verzekeringsartsen aan dat de belangrijkste belemmering voor gebruik was dat ze naar hun idee nog onvoldoende informatie over het instrument hadden. De resultaten van dit onderzoek laten zien dat het belangrijk is om verzekeringsartsen meer vertrouwd te maken met de inhoud en toepassing van voorspel-modellen en beslissingsondersteunende instrumenten om tot een succesvolle toepassing in de verzekeringsgeneeskundige praktijk te komen.

Discussie en aanbevelingen

In **hoofdstuk 7** zijn de belangrijkste bevindingen uit dit proefschrift samengevat en worden aanbevelingen gegeven voor onderzoek, beleid en praktijk. De resultaten van dit proefschrift laten zien dat voorspelmodellen en beslissingsondersteunende instrumenten kunnen helpen om te bepalen welke mensen een hoger risico hebben om langdurig een arbeidsongeschikt-heidsuitkering te ontvangen. Verzekeringsartsen en andere professionals kunnen het instrument raadplegen en als hulpmiddel gebruiken bij het maken van (wetenschappelijk) onderbouwde beslissingen rondom de prognose van werkvermogen, het doelmatig plannen van herbeoordelingen en de inzet van gerichte dienstverlening voor terugkeer naar werk. Voorspelmodellen moeten extern gevalideerd en geëvalueerd worden met effectmetingen en procesevaluaties om te bepalen wat hun toegevoegde waarde is voor de dagelijkse praktijk. Hiervoor moeten kleinschalige experimenten met dit soort modellen opgezet worden. Onderzoekers zijn hierbij afhankelijk van de bereidheid van artsen en andere professionals om mee te doen met dit soort experimenten. Omdat beslissingsondersteunende instrumenten bedoeld zijn als hulpmiddel, en niet om het werk van de professional over te nemen, is er bij deelname aan dit type experimenten geen sprake van beperking van de professionele autonomie. Professionals meer bekend maken met de toepassing en mogelijkheden van data analytics en voorspelmodellen, door deze onderwerpen bijvoorbeeld te behandelen tijdens de opleiding of trainingsdagen, is een belangrijke eerste stap.

About the author

Ilse Louwerse was born on 27th of January 1989 in Rotterdam, the Netherlands. She received both her bachelor's and master's degree in Econometrics and Management Science from Erasmus University Rotterdam. After her studies, she did a development internship at AIESEC and workerd as a junior researcher at the Erasmus University Rotterdam. Since 2015, she worked as customer intelligence analist at the Dutch Institute for Employee Benefit Schemes (UWV). In 2016, Ilse started her PhD at the Research Center for Insurance Medicine (KCVG), a joint initiative of the Academic Medical Center, the University Medical Center Groningen, UWV, and the VU University Medical Center (VUmc). She worked at the Department of Public and Occupational Health at Amsterdam UMC, location VUmc. Her research on use of predictive analytics in insurance medicine resulted in several international publications which form the basis of this thesis. Ilse currently works as a Senior Consultant Data & Analytics at The Next School.

List of publications

Articles included in this thesis

Louwerse I, Huysmans MA, van Rijssen HJ, van der Beek AJ, Anema JR. Characteristics of individuals receiving disability benefits in the Netherlands and predictors of leaving the disability benefit scheme: a retrospective cohort study with five-year follow-up. BMC Public Health. 2018;18(1):157.

Louwerse I, van Rijssen HJ, Huysmans MA, van der Beek AJ, Anema JR. Predicting long-term sickness absence and identifying subgroups among individuals without an employment contract. J Occup Rehabil. 2020;30(3):371-380.

Louwerse I, Huysmans MA, van Rijssen JH, Schaafsma FG, Weerdesteijn KH, van der Beek AJ, Anema JR. Predicting future changes in the work ability of individuals receiving a work disability benefit: weighted analysis of longitudinal data. Scand J Work Environ Health. 2020;46(2):168-176.

Louwerse I, Huysmans MA, van Rijssen JH, Overvliet J, van der Beek AJ, Anema JR. Preferences regarding the way of use and design of a work ability prognosis support tool: a focus group study among professionals. Disabil Rehabil. 2019 Nov 26:1-7.

Louwerse I, Huysmans MA, van Rijssen JH, Gielen CLI, van der Beek AJ, Anema JR. Use of a decision support tool on prognosis of work ability in work disability assessments: an experimental study among insurance physicians. J Occup Rehabil. 2020 Jun 11:1-12.

Other publications

Louwerse I, Huysmans MA, van Rijssen HJ, van der Beek AJ, Anema JR. Wie stroomt er in de WIA in en wie stroomt uit? Een vijf jaar follow-up studie. TBV jaargang 27, nr. 1, januari 2019.

Weerdesteijn KH, Schaafsma FG, **Louwerse I**, Huysmans MA, Van der Beek AJ, Anema JR. Does self-perceived health correlate with physician-assessed functional limitations in medical work disability assessments? J Psychosom Res. 2019 Aug 2:1-9.

Louwerse I, van Rijssen HJ, Huysmans MA, van der Beek AJ, Anema JR. Toekomstig werkvermogen van WIA-klanten voorspellen: Gewogen analyse van longitudinale data. TBV jaargang 28, nr. 3, maart 2020.

Sewdas R, de Wind A, Stenholm S, Coenen P, **Louwerse I**, van Boot CRL, van der Beek AJ. The association between retirement and mortality: working longer, living longer? – a systematic review and meta-analysis. J Epidemiol Community Health. 2020;74(5):473-480.

Louwerse I, Huysmans MA, van Rijssen HJ, Gielen CLI, van der Beek AJ, Anema JR. Gebruik van een beslishulp bij de prognosestelling. TBV jaargang 28, nr. 9, oktober 2020.

Dankwoord

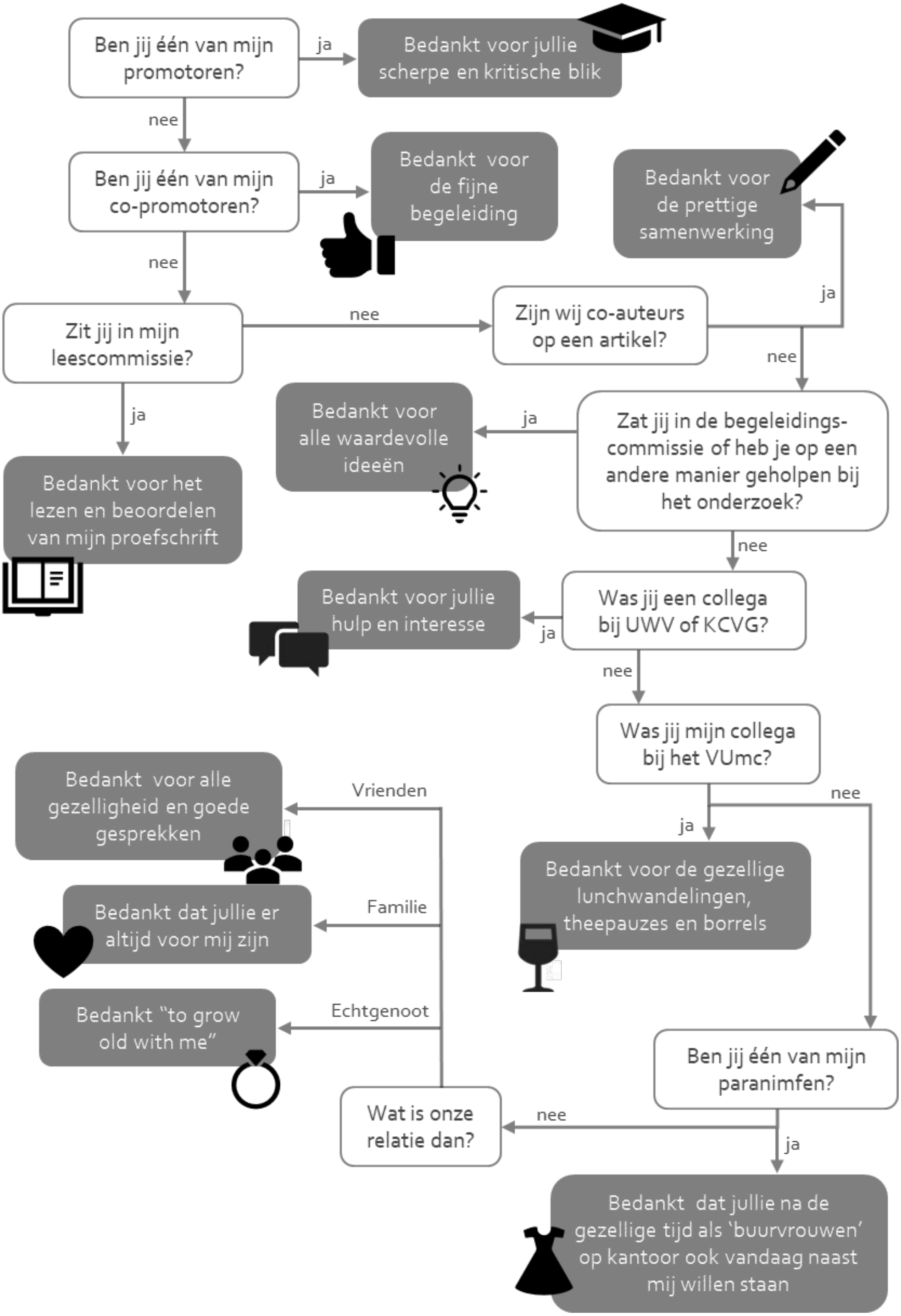
Als kind droomde ik ervan om dokter te worden. Ik stelde mezelf voor hoe het was om in een ziekenhuis te werken en mensen beter te maken. Helaas kwam ik er al snel achter dat ik bij het zien van naalden en bloed al flauwval, wat me geen handige eigenschap leek als ik een goede dokter wilde worden. Toen ik in 6 vwo zat, spoorde mijn vader mij aan om bij de studie econometrie te gaan kijken. Van die keuze heb ik geen moment spijt gehad (bedankt voor de tip pap!). Toch voelt het bijzonder dat ik de afgelopen vier jaar promotieonderzoek mocht doen bij het VUmc, en dat ik mij binnenkort doctor mag noemen!

Hoewel alleen mijn naam op de voorkant van dit boekje staat, heb ik dit onderzoek zeker niet alleen gedaan. Mijn grootste dank gaat uit naar mijn (co-)promotoren. Han, met mijn voorliefde voor data heb ik het je niet altijd makkelijk gemaakt. Bedankt dat je de afgelopen vier jaar toch promotor wilde zijn op dit project. Jij zorgde evoor dat ik het wetenschappelijke aspect van onderzoek doen nooit uit het oog verloor. Allard, tijdens onze teamoverleggen spoorde jij me aan om zelf knopen door te hakken. Jouw scherpe en duidelijke feedback prikkelde mij om mijn onderzoeksideeën verder aan te scherpen en helderder op te schrijven.

Maaike en Jolanda, ik heb het al vaker gezegd: jullie zijn een perfect duo. Qua persoonlijkheid vullen jullie elkaar goed aan en qua begeleiding zitten jullie (meestal) op een lijn. Jullie gaven mij veel vrijheid bij het doen van dit onderzoek, maar waren altijd betrokken. Maaike, wát ik ook wilde bespreken, jouw deur stond altijd open. Jij liet me stilstaan bij de successen maar voelde ook feilloos aan wanneer het even niet zo lekker liep. Dan gaf je mij het vertouwen om weer door te gaan. Jolanda, met jou had ik al op een hele fijne manier samengewerkt voor een eerder project bij UWV. Dat jij een van mijn co-promotoren zou worden heeft een belangrijke rol gespeeld in mijn keuze om dit promotieonderzoek te gaan doen. Ik had er alle vertouwen in dat ik met jouw begeleiding dit project succesvol af kon ronden en dat bleek helemaal terecht.

De praktische kant van mijn promotieonderzoek hadden we nooit kunnen doen zonder alle hulp die we vanuit UWV en de verschillende kantoren hebben gekregen. In het bijzonder wil ik Chantal bedanken voor haar bijdrage aan het opzetten en uitvoeren van de pilot en vignetten-studie. Bedankt voor al je praktische adviezen en de gezelligheid tijdens onze bezoekjes aan de verschillende kantoren.

Natuurlijk zijn er nog veel meer mensen zonder wie ik dit proefschrift nooit had kunnen schrijven. Dank voor jullie bijdragen, groot of klein, inhoudelijk of persoonlijk. Kijk op de volgende pagina waarvoor ik jou wil bedanken!


Ben jij één van mijn promotoren?
ja
Bedankt voor jullie scherpe en kritische blik
nee
Ben jij één van mijn co-promotoren?
ja
Bedankt voor de fijne begeleiding
nee
Bedankt voor de prettige samenwerking
ja
Zit jij in mijn leescommissie?
nee
Zijn wij co-auteurs op een artikel?
nee
ja
Bedankt voor het lezen en beoordelen van mijn proefschrift
Bedankt voor alle waardevolle ideeën
ja
Zat jij in de begeleidings-commissie of heb je op een andere manier geholpen bij het onderzoek?
nee
Bedankt voor jullie hulp en interesse
ja
Was jij een collega bij UWV of KCVG?
nee
Was jij mijn collega bij het VUmc?
nee
ja
Bedankt voor alle gezelligheid en goede gesprekken
Vrienden
Bedankt voor de gezellige lunchwandelingen, theepauzes en borrels
Bedankt dat jullie er altijd voor mij zijn
Familie
Bedankt "to grow old with me"
Echtgenoot
Ben jij één van mijn paranimfen?
Wat is onze relatie dan?
nee
ja
Bedankt dat jullie na de gezellige tijd als 'buurvrouwen' op kantoor ook vandaag naast mij willen staan