Quality improvement of vocational rehabilitation in patients with chronic musculoskeletal pain and reduced work participation

Timo Beemster

This thesis, titled “Quality improvement of vocational rehabilitation in patients with chronic musculoskeletal pain and reduced work participation” was a collaboration between:

- Department of Rehabilitation Medicine, Center for Rehabilitation, University Medical Center Groningen, University of Groningen;
- Department of Research and Development, Heliomare Rehabilitation Center, Wijk aan Zee, The Netherlands;
- Amsterdam UMC, University of Amsterdam, Coronel Institute of Occupational Health, Amsterdam Public Health research institute, Amsterdam, The Netherlands.

The thesis project was funded by Heliomare (seven years) and Amsterdam UMC - Coronel Institue of Occupational Health (one year). PhD training was facilitated by the Research Institute SHARE of the Graduate School for Medical Sciences and the AMC Graduate School.

The printing of this thesis was financially supported by:
University of Groningen - RUG
Graduate school for Health Research - SHARE
Centrum voor Revalidatie - Universitair Medisch Centrum Groningen (UMCG)
Revalidatiecentrum Heliomare - Research & Development
Amsterdam UMC - Coronel Instituut voor Arbeid en Gezondheid
Netwerk Vroege Interventie

Cover design: Richard Fickert
Layout: Marilou Maes, persoonlijkproefschrift.nl
Printing: Ridderprint BV | www.ridderprint.nl

Beemster, Timo. Quality improvement of vocational rehabilitation in patients with chronic musculoskeletal pain and reduced work participation. Thesis University of Groningen, the Netherlands – with references – with summary in Dutch.

ISBN: 978-94-034-1849-0 (printed version)
ISBN: 978-94-034-1848-3 (electronic version)

rijksuniversiteit groningen

Quality improvement of vocational rehabilitation in patients with chronic musculoskeletal pain and reduced work participation

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. E. Sterken
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 28 augustus 2019 om 16.15 uur

door

Timotheus Theodorus Beemster

geboren op 6 oktober 1985
te Hoorn

Promotores

Prof. dr. M.F. Reneman

Prof. dr. M.H.W. Frings-Dresen

Prof. dr. C.A.M. van Bennekom

Copromotor

Dr. J.M. van Velzen

Beoordelingscommissie

Prof. dr. R. Smeets

Prof. dr. P. Dijkstra

Prof. dr. S. Brouwer

Paranimfen

Jaap Beemster

Weijert Beemster

Table of content

CHAPTER 1

General Introduction

Background

Lower back and neck pain were the leading global causes of disability between 1990 and 2015 [1]. Between 2007 and 2011, 17% of the Dutch working population suffered from some form of musculoskeletal pain, such as low back, neck and shoulder pain [2]. Following the onset of musculoskeletal pain, most individuals (75-90%) recover within 8 weeks to become pain free [3,4]. However, 10-15% will be disabled at 3 months [4,5], representing a progression from acute to chronic musculoskeletal pain (CMP) [6]. The socioeconomic burden of CMP involves costs related directly to healthcare, loss of productivity, early retirement, and disability benefits [2, 7, 8]. In the Netherlands, 4.1 billion euros (3.7 percent of gross domestic product [9]) was spent in 2008 on disability and sickness benefits [10]; a significant proportion of this is directly related to patients with CMP. Furthermore, those with CMP are often unable to participate in work or full-time employment [2, 10]. Engaging in paid work has been proven to be of benefit at both a group and a patient level, providing income, enabling social relationships, structuring time, and supporting individual development [11, 12]. Therefore, achieving sustainable levels of work participation in workers with CMP is of significant importance from both a societal and individual perspective.

Vocational rehabilitation

Research has shown that multi-domain VR is beneficial in achieving sustainable levels of work participation in sick-listed workers with CMP [13-16]. VR can be understood as an interdisciplinary, multi-domain intervention program, comprising multimodal treatments provided by a multidisciplinary team, collaborating in the assessment and treatment of patients using a shared biopsychosocial model [17-21] and shared goals [22]. The primary aim of VR is to achieve and optimize work participation [23]. Secondary aims of VR might be the reduction of disability or health care usage. VR consists of components from three primary domains of intervention [16]:

1. Health-focused interventions, such as graded activity/physical exercises, cognitive behavioral therapy (CBT), education, and occupational therapy.
2. Service coordination interventions, such as the development of return to work (RTW) plans, case management, education, and training.
3. Work modification interventions, such as modified duties, modified working hours, supernumerary replacements (e.g., modified work), ergonomic adjustments, and other worksite adjustments.

A program is classed as “multi-domain” when it contains multiple intervention components from at least two of the three domains described [16].

Despite convincing evidence that VR can achieve sustainable work participation in patients with CMP [5, 14-16], there are a number of research gaps concerning the “clinimetrics” and “dose-content” of VR. This thesis looks explicitly at these factors—directly addressing deficiencies identified in the literature and supplementing the existing knowledge base—to the end of improving the overall quality of VR that can be delivered.

Clinimetrics research gap

In order to develop both clinical practice and research methodology, it is necessary to assess the clinimetric properties of VR. To do so, and to be able to relate this to VR effectiveness in our target population, we need to be able to assess the biopsychosocial characteristics of CMP populations, as well as measure the outcomes of interventions. No existing questionnaire set or measurement tool specifically tailored towards VR is currently available in the Netherlands; furthermore, the clinimetric properties of existing instruments focusing on work participation, healthcare usage, and disability are not directly applicable to the context of Dutch VR. These factors form part of the “clinimetrics research gap” identified in this thesis and will be addressed over the following four sections.

Core set development

For purposes of clinical practice and research, similar population characteristics and outcome measures (such as patient reported outcome measures, or “PROMS”) are collected to allow assessment of a specific clinical intervention. This enhances comparability (benchmarking) and allows researchers to develop studies in order to improve clinical and cost effectiveness [24]. Two measurement tools (questionnaires) have been developed in recent years that are directly relevant to the content of this thesis: one in the field of vocational rehabilitation (the brief ICF Core Set for vocational rehabilitation [25]) and one in the field of pain (the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT)) [26, 27]. These were developed for use in two broad clinical areas (pain and work); however, no *core set* of questionnaires exist that can address both factors simultaneously. In order to be able to adequately measure pain within the context of VR, a tool that can integrate both “pain” and “work” is required. The two aforementioned instruments can be useful in this context,

but they should be merged and made applicable to the specific context, patient group, and setting[1].

Measurement of work participation

The goal of multi-domain VR is to facilitate sustainable work participation [23]. Work participation can be quantified through the functions of absenteeism (referring to unscheduled employee absence from work) and presenteeism (referring to productivity loss while at work). These two constructs can be grouped as "productivity loss measures." Productivity loss can be estimated using nationwide disability benefit databases or using PROMS [16]. In contrast to other European countries, a nationwide sickness benefit database is not available to researchers within the Netherlands; hence, Dutch researchers must assess productivity loss through the use of PROMS. In order to identify and evaluate productivity loss rates before, during, or after VR, it is important that PROMS are derived in a way that is valid, reliable, and responsive. There have been several tools developed for looking at productivity loss over recent years [28-34], but these can not be directly applied to the Dutch VR context. Moreover, the reliability of existing productivity loss measures is, in general, poor [28, 32, 33]. Another shortcoming of existing tools is that information regarding *responsiveness* and *interpretation of change* of productivity loss measures is lacking. These measures are important to allow adequate evaluation of VR programs on both an individual and group level, and to enhance benchmarking.

Measurement of healthcare usage

Information concerning healthcare usage is required when performing cost-effectiveness analyses. A Dutch questionnaire, the Trimbos iMTA questionnaire measuring the costs of psychiatric illnesses (TiCP, part I), has been developed to assess healthcare usage in mental health patients [35]. This questionnaire showed adequate clinimetric properties and is recommended in the Dutch guideline for health economic evaluations [36]. The TiCP, however, is not directly applicable to the patient group and setting of this thesis, and should, therefore, be adapted and tested according to clinimetric principles.

Measurement of disability

Patients with CMP can suffer from many problems outside the workplace; for example, problems in self-care, childcare activities, and social participation

[1] In this thesis, this group refers to patients with CMP with reduced work participation, referred for VR in the Netherlands.

have all been described in relation to CMP. Therefore, an important secondary outcome of interdisciplinary VR is the level of *pain-related disability*. A valid and reliable questionnaire to measure pain-related disability is the Pain Disability Index (PDI) [37-40]. The PDI is a generic questionnaire and can, by definition, be applied to different patient groups, including those with chronic low back pain, fibromyalgia, cancer, or chronic widespread pain. The utility of the PDI is high because it is easy to understand and can be conducted over a short period, as it consists of only seven questions [41].

Nevertheless, a lack of consensus exists regarding how to interpret change scores in PDI following discharge from a treatment program. Information about the responsiveness of the PDI is needed to calculate the *interpretation of change score*. Responsiveness of the PDI has been previously studied in a Dutch pain rehabilitation setting [41], although this study did not account for measurement error. It is therefore unknown whether the cutoff point (i.e., the minimal important change) identified in this study represented real change or was affected by measurement error. Moreover, the change score for a multi-item questionnaire with a continuous outcome scale might vary according to baseline scores (they may be baseline dependent) [42-44]. It can be hypothesized, therefore, that patients with a high disability score at baseline should exhibit a greater increase in score on PDI—thus allowing us to infer that a clinically relevant change in pain-related disability has occurred—compared with patients with a low disability score at baseline. This hypothesis will be studied.

Dose-content research gap

As described previously, research has shown multi-domain VR to be beneficial in achieving sustainable levels of work participation in sick-listed workers with CMP [5, 14-16]. However, the effect sizes reported are moderate [14-16, 45]. Moreover, since existing programs are extensive and of high cost, there is a demand for simple, low-cost VR programs [13, 14]. It is unknown as to whether complex patient groups, such as those with CMP and reduced work participation, could benefit from such—simplified—programs. Application of VR programs also tends to be fairly nonspecific; it is therefore not fully understood which treatment components work best for whom. Optimal practice in the construction and application of VR programs have, in summation, not been comprehensively established. This is the second research gap addressed in this thesis, described henceforth as a “dose-content” issue, and is explored over the next two sections.

Content

A wide range of content exists across the VR programs described in the literature [14-16, 45,46] both in terms of their constituent components and their domains of application. This hinders guideline development and the development of specific recommendations for rehabilitation centers and policymakers. A systematic review [14] of the effectiveness of community- and workplace-based interventions in musculoskeletal-related sickness absence showed that concerning the earlier mentioned intervention domains (health-focused, service coordination, work modification), five studies (12%) contained components from all three domains, 12 studies (29%) contained components from the "health-focused" and either "work modification" or "service coordination" domains, 21 studies (50%) contained components only from the "health-focused" domain, and 4 studies (10%) contained only components from the "work modification" or "service coordination" domain [14]. The review advised a focus on the implementation of simple, low-cost interventions containing a work-based or primary care element, as these interventions are the most feasible to conduct in clinical or workplace practice, and might be the most cost effective [14]. A disadvantage of the review methodology, however, was that the majority of included studies were conducted in subacute musculoskeletal pain patients. It is unknown, therefore, as to whether their conclusions are applicable to patients with chronic musculoskeletal pain.

Another review [16], aiming to explore the effectiveness of workplace interventions on work participation in musculoskeletal, pain-related, and mental health conditions, showed similar patterns in treatment program content. This review identified 15 (42%) single-domain studies and 21 (58%) multi-domain studies. Of the latter, 15 studies contained treatment components from all three domains. The authors concluded that multi-domain interventions, with components from at least two of the three domains, can help reduce time lost from work in CMP-related conditions [16]. These two review articles present contrasting conclusions and recommendations for VR program design: should a VR program be *comprehensive* (consisting of multiple components from all three domains), or simpler, and *less comprehensive* (containing fewer components from two domains)? Given this lack of consensus, it is meaningful to explore the core components of clinically- and cost-effective multi-domain VR.

Dosage

It is currently unknown as to what *dosage* of VR treatment (a term incorporating treatment duration, intensity, number of contact hours, and number of disciplines

involved) is optimal for patients with CMP and reduced work participation. The literature describes a variety of differing dosages between programs with, apparently, little impact on work participation levels. For example, a review showed that effective VR programs for patients with CMP ranged from those with 6 contact hours to those containing more than 70 contact hours [14]. Another review demonstrated that pain rehabilitation programs containing 7 to 197 contact hours were effective in enhancing the work participation of patients with CMP [45]. Furthermore, the last two decades have provided a growing evidence base for the premise that less-comprehensive vocational rehabilitation programs may be non-inferior (when compared to comprehensive programs) in their impact on work participation [45, 47-53]. For instance, several randomized controlled trials have shown that VR programs with differing numbers of contact hours were non-inferior to each other with regard to enhancing the work participation of sick-listed workers with CMP (e.g., 18.5 hrs vs. 52 hrs [47], 15 hrs vs. 120 hrs [52], and 10 hrs vs. 120 hrs [54, 55]). In addition, a Dutch qualitative study showed that patients' and clinicians' satisfaction with a pain rehabilitation program was independent of the program dosage [56]. Thus far, no quantitative "dose-response" studies have been performed in the Netherlands. As VR programs in the Netherlands are commonplace, and since evidence has shown that geographic location can affect rehabilitation results [57], a dose-response study looking specifically at VR in the Netherlands can be justified.

Thesis objective and research questions

The overall aim of this thesis is to contribute to the quality improvement of vocational rehabilitation for patients with chronic musculoskeletal pain and reduced work participation.

The aim of this thesis is divided into two parts:

I. To investigate the clinimetric properties of work participation, healthcare usage, and pain-related disability measures.
II. To investigate the relationship between the dosage and content of VR on work participation.

The research questions of this thesis are:

Part I: Clinimetric

1. Which questionnaires should be included in a focused "VR-pain Core Set" that can be used across VR practice in the Netherlands and can examine clinical and cost effectiveness?
2. What are the clinimetric properties of work participation, healthcare usage, and pain-related disability questionnaires for patients with CMP and reduced work participation in attendance of, and following discharge from, VR in the Netherlands?

Part II: Dose-content

3. What are the opinions and experiences of patients, professionals, and managers regarding the usefulness and feasibility of "comprehensive" and "less-comprehensive" VR programs?
4. Are patients with CMP and reduced work participation who attended "VR with work module" more likely to achieve work participation than patients who attended "VR without work module?"

Thesis outline

Part I: Clinimetric. Research questions 1 and 2 are answered with Chapters 2-4.

- **Chapter 2**: development of a consensus-based "VR-pain Core Set" of patient-reported outcome measures for use in patients with CMP and reduced work participation enrolled in VR programs in the Netherlands.

- **Chapter 3**: examination of the reliability, agreement, and responsiveness of a work productivity questionnaire (iPCQ-VR) and a healthcare usage questionnaire (TiCP-VR), both developed for patients with CMP and reduced work participation in attendance of, and following discharge from, VR in the Netherlands.

- **Chapter 4**: determination of the *responsiveness* and *interpretation of change scores* of the Pain Disability Index, in patients with CMP and reduced work participation at discharge from VR.

Part II: Dose-content. Research questions 3 and 4 will be answered with Chapters 5-7.

- **Chapter 5**: a study protocol for a multicenter, randomized controlled trial aiming to study the effectiveness and cost effectiveness of "comprehensive" and "less-comprehensive" VR in patients with CMP and reduced work participation.

- **Chapter 6**: a qualitative study, in which patients, professionals, and managers with experiences in a multicenter RCT (Chapter 5) were asked about the usefulness and feasibility of "comprehensive" and "less-comprehensive" VR programs.

- **Chapter 7**: a retrospective cohort study, in which the likelihood of successful work participation following a VR program with or without work module was assessed.

Chapter 8 is the general discussion of this thesis.

References

1. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016 Oct 8;388(10053):1545-602.
2. de Vroome EM, Uegaki K, van der Ploeg CP, Treutlein DB, Steenbeek R, de Weerd M, et al. Burden of Sickness Absence Due to Chronic Disease in the Dutch Workforce from 2007 to 2011. J Occup Rehabil. 2015 Dec;25(4):675-84.
3. Hayden JA, Dunn KM, van der Windt DA, Shaw WS. What is the prognosis of back pain? Best Pract Res Clin Rheumatol. 2010 Apr;24(2):167-79.
4. Oude Nijeweme–d'Hollosy W, Van Velsen L, Soer R, Hermens H. Design of a Web-based Clinical Decision Support System for Guiding Patients with Low Back Pain to the Best Next Step in Primary Healthcare. The 9th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2016) 2016. p. 229-39.
5. Waddell G, Burton A, Kendall N. Vocational rehabilitation: what works, for whom, and when? The Stationery Office 2013.
6. Steenstra IA, Munhall C, Irvin E, Oranye N, Passmore S, Van Eerd D, et al. Systematic Review of Prognostic Factors for Return to Work in Workers with Sub Acute and Chronic Low Back Pain. J Occup Rehabil. 2017 Sep;27(3):369-81.
7. Lambeek LC, van Tulder MW, Swinkels IC, Koppes LL, Anema JR, van Mechelen W. The trend in total cost of back pain in The Netherlands in the period 2002 to 2007. Spine (Phila Pa 1976). 2011 Jun;36(13):1050-8.
8. Escorpizo R HD, Stucki G, Brage S. Handbook of Vocational Rehabilitation and Disability Evaluation. Springer, 2014.
9. Sickness, Disability and Work: Breaking the Barriers. A synthesis of findings across OECD countries. OECD 2010.
10. Steenbeek R, Hooftman W, Geuskens G, Wevers C. Objectiveren van gezondheidsgerelateerde nonparticipatie en de vermijdbare bijdrage van de gezondheidszorg hieraan (Objectification of health-related non-participation and the avoidable contribution of health care to this) TNO 2010.
11. Sheppard DM, Frost D. A new vocational rehabilitation service delivery model addressing long-term sickness absence. British Journal of Occupational Therapy. 2016;79(11):677-81
12. Saunders SL, Nedelec B. What work means to people with work disability: a scoping review. J Occup Rehabil. 2014 Mar;24(1):100-10.
13. Aasdahl L, Fimland MS. Is there really a "golden hour" for work disability interventions? a narrative review. Disabil Rehabil. 2019 Jan;(3):1-8.
14. Palmer KT, Harris EC, Linaker C, Barker M, Lawrence W, Cooper C, et al. Effectiveness of community- and workplace-based interventions to manage musculoskeletal-related sickness absence and job loss: a systematic review. Rheumatology (Oxford). 2012 Feb;51(2):230-42.
15. van Vilsteren M, van Oostrom SH, de Vet HC, Franche RL, Boot CR, Anema JR. Workplace interventions to prevent work disability in workers on sick leave. Cochrane Database Syst Rev. 2015 Oct 5(10):CD006955.

16. Cullen KL, Irvin E, Collie A, Clay F, Gensby U, Jennings PA, et al. Effectiveness of Workplace Interventions in Return-to-Work for Musculoskeletal, Pain-Related and Mental Health Conditions: An Update of the Evidence and Messages for Practitioners. J Occup Rehabil. 2018 Mar;28(1):1-15.
17. Engel GL. The need for a new medical model: a challenge for biomedicine. Science. 1977 Apr 8;196(4286):129-36.
18. Waddel G, Burton AK. Concepts of rehabilitation for the management of common health problems. The Stationery Office 2004.
19. Wade D. Rehabilitation--a new approach. Overview and Part One: the problems. Clin Rehabil. 2015 Nov;29(11):1041-50.
20. Wade D. Rehabilitation - a new approach. Part two: the underlying theories. Clin Rehabil. 2015 Dec;29(12):1145-54.
21. Wade D. Rehabilitation - a new approach. Part three: the implications of the theories. Clin Rehabil. 2016 Jan;30(1):3-10.
22. Task Force on Multimodal Pain Treatment Defines Terms for Chronic Pain Care. Accessed 22 March 2019; Available from: http://www.iasp-pain.org/PublicationsNews/NewsDetail.aspx?ItemNumber=6981.
23. Escorpizo R, Reneman MF, Ekholm J, Fritz J, Krupa T, Marnetoft SU, et al. A conceptual definition of vocational rehabilitation based on the ICF: building a shared global model. J Occup Rehabil. 2011 Jun;21(2):126-33.
24. Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for patients. Trials. 2017 Mar 14;18(1):122.
25. Finger ME, Escorpizo R, Glassel A, Gmunder HP, Luckenkemper M, Chan C, et al. ICF Core Set for vocational rehabilitation: results of an international consensus conference. Disabil Rehabil. 2012;34(5):429-38.
26. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain. 2003 Dec;106(3):337-45.
27. Turk DC, Dworkin RH, Revicki D, Harding G, Burke LB, Cella D, et al. Identifying important outcome domains for chronic pain clinical trials: an IMMPACT survey of people with pain. Pain. 2008 Jul 15;137(2):276-85.
28. Beaton DE, Tang K, Gignac MA, Lacaille D, Badley EM, Anis AH, et al. Reliability, validity, and responsiveness of five at-work productivity measures in patients with rheumatoid arthritis or osteoarthritis. Arthritis Care Res (Hoboken). 2010 Jan;62(1):28-37.
29. Bouwmans C, Krol M, Severens H, Koopmanschap M, Brouwer W, Hakkaart-van Roijen L. The iMTA Productivity Cost Questionnaire: A Standardized Instrument for Measuring and Valuing Health-Related Productivity Losses. Value Health. 2015 Sep;18(6):753-8.
30. Krol M, Brouwer W. How to estimate productivity costs in economic evaluations. Pharmacoeconomics. 2014 Apr;32(4):335-44.
31. Krol M, Brouwer W, Rutten F. Productivity costs in economic evaluations: past, present, future. Pharmacoeconomics. 2013 Jul;31(7):537-49.

32. Leggett S, van der Zee-Neuen A, Boonen A, Beaton DE, Bojinca M, Bosworth A, et al. Test-retest Reliability and Correlations of 5 Global Measures Addressing At-work Productivity Loss in Patients with Rheumatic Diseases. J Rheumatol. 2016 Feb;43(2):433-9.

33. Tang K. Estimating productivity costs in health economic evaluations: a review of instruments and psychometric evidence. Pharmacoeconomics. 2015 Jan;33(1):31-48.

34. Zhang W, Bansback N, Anis AH. Measuring and valuing productivity loss due to poor health: A critical review. Soc Sci Med. 2011 Jan;72(2):185-92.

35. Bouwmans C, De Jong K, Timman R, Zijlstra-Vlasveld M, Van der Feltz-Cornelis C, Tan Swan S, et al. Feasibility, reliability and validity of a questionnaire on healthcare consumption and productivity loss in patients with a psychiatric disorder (TiC-P). BMC Health Serv Res. 2013;13:217.

36. Hakkaart-van Roijen L, Tan S, Bouwmans C. Cost research guide. Methods and standard cost prices for economic evaluations in health care. [In Dutch: Handleiding voor kostenonderzoek. Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg]. Rotterdam: College voor zorgverzekeringen 2010.

37. Chibnall JT, Tait RC. The Pain Disability Index: factor structure and normative data. Arch Phys Med Rehabil. 1994 Oct;75(10):1082-6.

38. Gronblad M, Hupli M, Wennerstrand P, Jarvinen E, Lukinmaa A, Kouri JP, et al. Intercorrelation and test-retest reliability of the Pain Disability Index (PDI) and the Oswestry Disability Questionnaire (ODQ) and their correlation with pain intensity in low back pain patients. Clin J Pain. 1993 Sep;9(3):189-95.

39. Jerome A, Gross RT. Pain disability index: construct and discriminant validity. Arch Phys Med Rehabil. 1991 Oct;72(11):920-2.

40. Soer R, Koke AJ, Vroomen PC, Stegeman P, Smeets RJ, Coppes MH, et al. Extensive validation of the pain disability index in 3 groups of patients with musculoskeletal pain. Spine (Phila Pa 1976). 2013 Apr 20;38(9):E562-8.

41. Soer R, Reneman MF, Vroomen PC, Stegeman P, Coppes MH. Responsiveness and minimal clinically important change of the Pain Disability Index in patients with chronic back pain. Spine (Phila Pa 1976). 2012 Apr 15;37(8):711-5.

42. Baker DW, Hays RD, Brook RH. Understanding changes in health status. Is the floor phenomenon merely the last step of the staircase? Med Care. 1997 Jan;35(1):1-15.

43. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Ann Rheum Dis. 2005 Jan;64(1):29-33.

44. Olsen MF, Bjerre E, Hansen MD, Tendal B, Hilden J, Hrobjartsson A. Minimum clinically important differences in chronic pain vary considerably by baseline pain and methodological factors: systematic review of empirical studies. J Clin Epidemiol. 2018 Sep;101:87-106.

45. Waterschoot FP, Dijkstra PU, Geertzen JH, Reneman MF. Dose or content? Effectiveness of pain rehabilitation programs for patients with chronic low back pain: A systematic review. Pain. 2014 Jan;155(1):179-89.

46. Kamper SJ, Apeldoorn AT, Chiarotto A, Smeets RJ, Ostelo RW, Guzman J, et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain: Cochrane systematic review and meta-analysis. BMJ. 2015 Feb 18;350:h444.

47. Aasdahl L, Pape K, Vasseljen O, Johnsen R, Gismervik S, Halsteinli V, et al. Effect of Inpatient Multicomponent Occupational Rehabilitation Versus Less Comprehensive Outpatient Rehabilitation on Sickness Absence in Persons with Musculoskeletal- or Mental Health Disorders: A Randomized Clinical Trial. J Occup Rehabil. 2017 Apr 11.

48. Brendbekken R, Eriksen HR, Grasdal A, Harris A, Hagen EM, Tangen T. Return to Work in Patients with Chronic Musculoskeletal Pain: Multidisciplinary Intervention Versus Brief Intervention: A Randomized Clinical Trial. J Occup Rehabil. 2017 Mar;27(1):82-91.

49. Jensen C, Jensen OK, Nielsen CV. Sustainability of return to work in sick-listed employees with low-back pain. Two-year follow-up in a randomized clinical trial comparing multidisciplinary and brief intervention. BMC Musculoskelet Disord. 2012 Aug 25;13:156.

50. Moll LT, Jensen OK, Schiottz-Christensen B, Stapelfeldt CM, Christiansen DH, Nielsen CV, et al. Return to Work in Employees on Sick Leave due to Neck or Shoulder Pain: A Randomized Clinical Trial Comparing Multidisciplinary and Brief Intervention with One-Year Register-Based Follow-Up. J Occup Rehabil. 2018 Jun;28(2):346-56.

51. Myhre K, Marchand GH, Leivseth G, Keller A, Bautz-Holter E, Sandvik L, et al. The effect of work-focused rehabilitation among patients with neck and back pain: a randomized controlled trial. Spine (Phila Pa 1976). 2014 Nov 15;39(24):1999-2006.

52. Ronzi Y, Roche-Leboucher G, Begue C, Dubus V, Bontoux L, Roquelaure Y, et al. Efficiency of three treatment strategies on occupational and quality of life impairments for chronic low back pain patients: is the multidisciplinary approach the key feature to success? Clin Rehabil. 2017 Oct;31(10):1364-73.

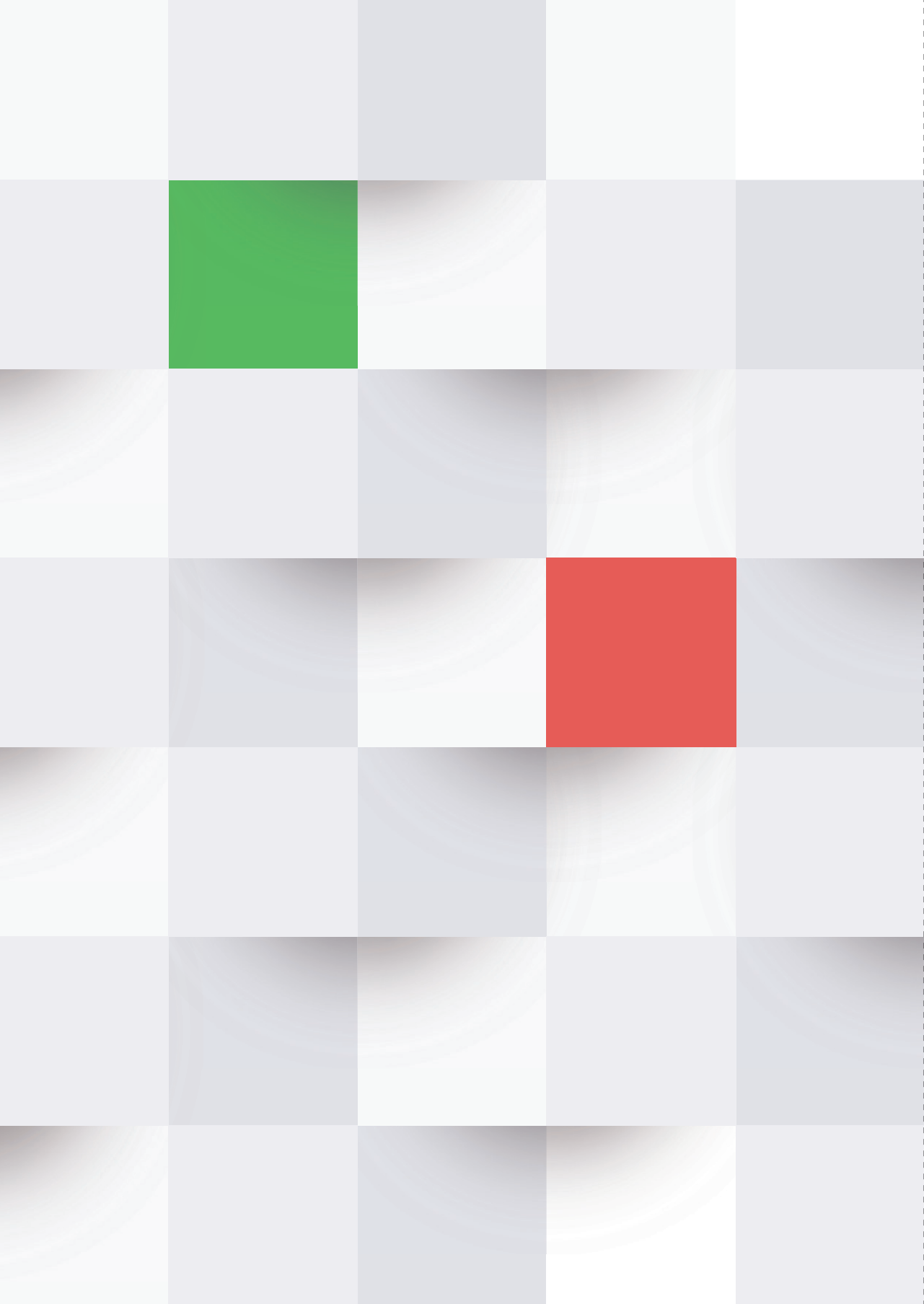
53. Reme SE, Tveito TH, Harris A, Lie SA, Grasdal A, Indahl A, et al. Cognitive Interventions and Nutritional Supplements (The CINS Trial): A Randomized Controlled, Multicenter Trial Comparing a Brief Intervention With Additional Cognitive Behavioral Therapy, Seal Oil, and Soy Oil for Sick-Listed Low Back Pain Patients. Spine (Phila Pa 1976). 2016 Oct 15;41(20):1557-64.

54. Haldorsen EM, Grasdal AL, Skouen JS, Risa AE, Kronholm K, Ursin H. Is there a right treatment for a particular patient group? Comparison of ordinary treatment, light multidisciplinary treatment, and extensive multidisciplinary treatment for long-term sick-listed employees with musculoskeletal pain. Pain. 2002 Jan;95(1-2):49-63.

55. Skouen JS, Grasdal AL, Haldorsen EM, Ursin H. Relative cost-effectiveness of extensive and light multidisciplinary treatment programs versus treatment as usual for patients with chronic low back pain on long-term sick leave: randomized controlled study. Spine (Phila Pa 1976). 2002 May 1;27(9):901-9.

56. Reneman MF, Waterschoot FPC, Bennen E, Schiphorst Preuper HR, Dijkstra PU, Geertzen JHB. Dosage of pain rehabilitation programs: a qualitative study from patient and professionals' perspectives. BMC Musculoskelet Disord. 2018 Jun 30;19(1):206.

57. Anema JR, Schellart AJ, Cassidy JD, Loisel P, Veerman TJ, van der Beek AJ. 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 Dec;19(4):419-26.



CHAPTER 2

Towards an ICF- and IMMPACT-based Pain Vocational Rehabilitation Core Set in the Netherlands

M. Reneman
T. Beemster
M. Edelaar
J. van Velzen
C. van Bennekom
R. Escorpizo

J Occup Rehabil 2013 Dec;23(4):576-84

Abstract

Background

For clinical use and research of pain within the context of vocational rehabilitation, a specific core set of measurements is needed. The recommendations of the International Classification of Functioning, Disability and Health (ICF) brief Core Set for Vocational Rehabilitation (VR) and those of Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) cover two broad areas. These two sources can be integrated when made applicable to vocational rehabilitation and pain.

Objective

To develop a core set of diagnostic and evaluative measures specifically for vocational rehabilitation of patients with subacute and chronic musculoskeletal pain, while using the brief ICF core set for VR as the reference framework in VR, and the IMMPACT recommendations in the outcome measurements around pain.

Methods

Three main steps were taken. The first step was to remove irrelevant and duplicate domains of the brief ICF Core Set for Vocational Rehabilitation and the IMMPACT recommendations around pain. The second step was to match the remaining domains with existing instruments or measures. Instruments were proposed based on availability and its proven use in Dutch practice and based on proof of sufficient clinimetric properties. In step 3, the preliminary VR-Pain core set was presented to 3 expert panels: proposed users, Dutch pain rehabilitation experts, and international VR experts.

Results

Experts agreed with the majority of the proposed domains and instruments. The final VR-Pain Core Set consists of 18 domains measured with 12 instruments. All instruments possessed basic clinimetric properties.

Conclusion

An agreed-upon VR-Pain Core Set with content that covers relevant domains for pain and VR and validated instruments measuring these domains has been developed. The VR-Pain Core Set may be used for regular clinical

purposes and research in the field of vocational rehabilitation and pain, but adaptations should be considered for use outside the Netherlands.

Keywords
ICF, IMMPACT, musculoskeletal pain, vocational rehabilitation, work rehabilitation, employment, return to work.

Introduction

Chronic musculoskeletal pain has a substantial negative impact on quality of life and the ability to engage in meaningful activities and participation in the society, including work [1, 2]. In the Netherlands, musculoskeletal disorders such as back, neck and shoulder pain constitute about 35% of all sickness absence and long-term disability compensations [3-5]. Medical care utilization and sickness absence due to musculoskeletal pain are associated with high economic burden to society similar to other western countries worldwide [6]. The majority (~80%) of the costs are related to the inability to work [5]. One of the preferred interventions to promote return to work for patients with chronic musculoskeletal pain is vocational rehabilitation, because it has been proven to be effective in reducing disability and improving work participation, and it appears to be cost-effective [7, 8]. To further improve the effectiveness of vocational rehabilitation to optimize work participation, it is recommended to intervene as soon as possible, perhaps even as soon as the sub-acute phase of musculoskeletal pain [9, 10]. A network of 14 rehabilitation centers in the Netherlands has been established to deliver evidence-based vocational rehabilitation for workers with sub-acute and chronic musculoskeletal pain.

Vocational rehabilitation (VR) in its broadest form has recently been defined in a position paper as 'a multidisciplinary evidence-based approach that is provided along a continuum of services and activities to working age individuals with health-related impairments, limitations, or restrictions with work functioning, and whose primary aim is to optimize work participation [11]. The authors of the position paper proposed the use of the International Classification of Functioning, Disability and Health (ICF) within the VR field (regardless of health condition). On one hand, the ICF Core Set for Vocational Rehabilitation has been developed with two versions: 1) the 90 ICF categories of the comprehensive version is

intended for multidisciplinary setting and 2) the 13 ICF categories of the brief version is intended for single discipline encounter or clinical trials. The brief version due to less number of ICF categories is doable for practical application and feasible in VR-related patient evaluation and assessment [12]. On the other hand, the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT), has provided recommendations for improving the design, execution, and interpretation of clinical trials of treatments specific to patients with pain [13, 14]. IMMPACT has proposed a core set of outcome measures for patients with pain. These recommendations were also broad, because they were intended to encompass the wide field of pain care, which extends far beyond the context of vocational rehabilitation.

So, here are two sets of recommendations or sets of domains, ICF Core Set and IMMPACT, which broadly address VR and pain, respectively. For clinical use and research of pain within the context of VR, developing a specific core set is needed, while learning from the two existing sets. The recommendations of the brief ICF Core Set for Vocational Rehabilitation and those of IMMPACT cover two broad areas, but should be merged and made applicable to a specific context, patient group and setting (pain and VR in the Netherlands in this study). However, we did not find papers relevant to the Netherlands, or anywhere else, describing the process and outcome of an ICF-IMMPACT core set, let alone the operationalization of those domains. The aim of the present study was to develop a core set of diagnostic and evaluative (clinical and economic) measures specifically for vocational rehabilitation of patients with sub-acute and chronic musculoskeletal pain, while using the ICF as the reference framework in VR, and IMMPACT in the outcome measurements around pain. In this study, the context is situated in the Netherlands, including its health care and social security policies as of the year 2012. As part of integrating our knowledge on the ICF, work, and pain, our research question is: how can the brief version of the ICF Core Set for Vocational Rehabilitation and the IMMPACT recommendations be best applied in one blended VR-Pain core set for patients with sub-acute and chronic musculoskeletal pain in the Netherlands?

Methods

Three main steps were taken. The first step was to remove irrelevant and duplicate domains of the brief ICF Core Set for Vocational Rehabilitation and the IMMPACT recommendations. Irrelevant domains were defined as those domains that do not apply or only apply to an estimated 1% of the target population (as judged by the authors and the expert panel). Duplicate domains were defined as domains that cover overlapping, equal or very similar content or concept. Additionally, the remaining domains were checked to see whether they could be used for economic evaluations also. If not, this was added. The second step was to match the remaining domains with existing instruments or measures. Instruments were proposed based on availability and its proven use in Dutch practice and peer reviewed literature. Existing instruments were included based on proof of sufficient reliability (test-retest reliability: Intra Class Coefficient (ICC) >0.90 (preferred), Kappa >0.60, Pearson correlation coefficient >0.80; internal consistency: Cronbach's alpha >0.80 [15]; construct validity (yes/no/not applicable (na)); responsiveness to change (yes/no/na; relevant for outcome measures only); existence of a validated version in Dutch language (yes/no; relevant for questionnaires only); and feasibility (acceptable patient and practitioner burden: yes/no). The second step was not performed to provide a systematic review of the psychometric properties of all instruments available, but to check whether the psychometric properties of the proposed instruments of the preliminary VR-Pain Core Set were acceptable.

The result of step 1 and 2 was a preliminary version of what we would call the VR-Pain core set. In step 3, to be informed by input from relevant people, the preliminary VR-Pain core set was presented to 3 expert or user panels: Dutch VR centers (proposed users (management and clinicians); n=13), Dutch pain rehabilitation development centers (pain rehabilitation experts; n=4), and members of the VR-Pain Core Set consensus group (VR experts; n=23) [12]. Participants were sent the introduction to, methods and results of steps 1 and 2, including the preliminary VR-Pain core set. They were asked whether they agreed with the taken steps and the proposed core domains of the preliminary VR-Pain Core Set, and whether they agreed with the proposed instruments. In case of non-agreement, they were asked to explain their disagreement and to suggest improvements. In case the comments were unclear, the first author contacted the responder. All participants had 3 weeks to respond. Participants were sent a reminder after 2 weeks. The authors of this paper then synthesized the comments of the responders into a final VR-Pain core set.

Results

Step 1.
The domains of the brief VR-Pain Core Set and the IMMPACT recommendations are presented in Tables 1 and 2. The results of Step 1, the selection of irrelevant domains and reduction of duplicates, are also presented in Tables 1 and 2.

Step 2.
Results of the process of matching core set domains to instruments, including its quality appraisal, are presented in Table 3. Additions as described in step 3 were also incorporated in Table 3. Domains from the IMMPACT recommendations were provided with ICF codes, with exception of personal factors which are not currently coded in the ICF.

Step 3.
The preliminary VR-Pain core set was emailed to members of the expert panels in February 2012. Overall response was n=18 (response rate 45%); proposed users n=11 (85%), pain rehabilitation experts n=4 (100%), VR experts n=3 (13%). Of the VR experts, an additional n=3 responded that the specific nature of the subject of this study was out of their field of expertise. One of the VR experts was contacted by phone, because the answers and comments were ambiguous. Eleven (61%) respondents agreed with the proposed domains of the preliminary core set, while five disagreed, and two did not answer. Ten (55%) respondents agreed with the proposed instruments of the preliminary core set, five disagreed, and three did not answer or indicated to have insufficient knowledge to judge. 'Disagreements' were most often accompanied by a short explanation and/or suggestion. The project members have decided unanimously that some comments should not be regarded as disagreements with the proposed domains or instruments, but rather as an item that a single expert proposed to add to the preliminary set. However, because not single experts, but rather the brief ICF Core Set for Vocational Rehabilitation and the IMMPACT recommendations formed the basis of this new and specific core set, it was decided that items proposed by single experts were not added to the definitive set, unless the project team decided otherwise based on the underlying core sets.

Based on the responses of the participants, the following domains were added to the VR-Pain Core set: adverse effects that has not lead to discontinuation of the program (adherence to the intervention; treatment records) and personal

Table 1. ICF categories of the brief ICF Core Set for Vocational Rehabilitation and relevance of the domains to the proposed VR-Pain core set

ICF Code	ICF Category Title	Relevant	Comments
Activities & Participation			
d155	Acquiring skills	No	This is not a key challenge in patients with pain. This item was included in the brief VR-Pain Core Set to accommodate individuals with neurological diagnoses and intellectual and cognitive challenges.
d240	Handling stress and other psychological demands	Yes	
d720	Complex interpersonal interactions	Yes	
d845	Acquiring, keeping and terminating a job	No	The target population is employed. Aim of VR in our case is to return to own work and same employer, or to improve work performance. Keeping a job: duplicate concept with d850.
d850	Remunerative employment	Yes	Work status will be assessed, including absenteeism and presenteeism.
d855	Non-remunerative employment	No	Only patients with paid work are admitted to our specified setting.
Environmental Factors			
e310	Immediate family	Yes	
e330	People in positions of authority	Yes	
e580	Health services, systems and policies	No	Within the target population, this item is of relevance, but not variable across subjects in the Netherlands.
e590	Labour and employment services, systems and policies	No	Within the target population, this item is of relevance, but not variable across subjects in the Netherlands.
Body Functions			
b130	Energy and drive functions	Yes	
b164	Higher-level cognitive functions	No	Within this target population, high-level cognitive functions are unaffected.
b455	Exercise tolerance functions	Yes	

problems unrelated to work (Work Reintegration Questionnaire; WRQ). With regard to the instruments, the following measurements were changed or added: energy and drive functions (ICF code b130) will be measured with numerical rating scale (NRS) for fatigue; physical functioning will not be measured with the Pain Disability Index only, but also with RAND-36 scale physical functioning; Astrand or Bruce submaximal ergometry will be used to measure exercise tolerance functions; assessment of functioning at home or in unpaid work will be added as part of the demographic questionnaire.

Description of Instruments of the final VR-Pain core set
The *EuroQol-5D* (EQ-5D) is a 6-item questionnaire to investigate quality of life. The EQ-5D categories measure 5 dimensions: mobility, self-care, activities of daily life, pain and anxiety/depression. Five questions are categorical (1-3 scale) and one question assessing general health status is on interval level (VAS 0-100). A Dutch language version of the EQ-5D is available [16, 17]. The EQ-5D is a widely employed instrument to assess health related quality of life (QoL), is used in cost effectiveness research based on Quality Adjusted Life Years (QALY) and is recommended by the Dutch Healthcare Insurance Board [18]. Lower levels of QoL are associated with productivity loss in patients with low back pain [19].

A single item of the *Work Ability Index* (WAI) will be used to assess self-reported work ability. Current work ability compared to lifetime best can be scored on a 0-10 response scale, where 0 represents 'completely unable to work' and 10 'work ability at its best'. A strong association between the single item and the complete WAI was observed (r=0.87) [20].

The *PROductivity and DISease Questionnaire* (PRODISQ) [21] will be used to assess employment status, absenteeism and presenteeism. Absenteeism refers to time off from work. Presenteeism refers to productivity loss while at-work. Both may be associated with a health condition. Absenteeism is measured with a three-month recall period, and will be measured specifically related to pain condition. The number and duration of a maximum number of three absenteeism periods are collected. Presenteeism is measured with two items on a 11-point scale, also known as the QQ-index (quantity and quality). The first item measures quality of work done in the last day at work, ranging from 0 (I couldn't do anything) to 10 (I could do the same as normal). The second item measures quantity of work done in the last day at work, ranging from 0 (the quality of my work was dramatic) to 10 (the quality of my work was normal).

Table 2. IMMPACT recommendations and supplemental domains and relevance of the domains to the proposed VR-Pain core set

	Relevant	Comments
Core domains		
Pain	Yes	
Physical functioning	Yes	
Emotional functioning	Yes	
Participant ratings of global improvement	Yes	
Symptoms and adverse events	Yes	Symptoms duplicate with pain. Adverse events will be monitored under participant disposition.
Participant disposition (including adherence to the treatment regimen and reasons for premature withdrawal from the trial)	Yes	Will be replaced by: Adherence to the intervention and reasons for premature withdrawal.
Supplemental domains		
Role functioning (i.e. work and educational activities)	Yes	
Interpersonal functioning (i.e. relationships and activities with family, friends, and others)	Yes	Duplicate. Will be covered under immediate family and people in authority (as mentioned in the ICF-VR), which are the primary group of interest in our context
Pharmacoeconomic measures and health care utilization	Yes	Will be included as one domain: health care utilization.
Biological markers (e.g. assessments based on quantitative sensory testing, imaging, genetic markers, pharmacogenomics, and punch skin biopsy)	No	The target population includes patients with non-specific pain. If biological functions are relevantly involved in the health status, patients are excluded because this could indicate a specific pain syndrome.
Coping	Yes	
Clinician or surrogate ratings of global improvement	Yes	
Neuropsychological assessments of cognitive and motor function	Yes	Duplicate. Will be covered under coping / stress and psychological demands and exercise tolerance and physical functioning, all part of ICF-VR
Suffering and other end-of-life issues	No	Not applicable for the target population.

The *Pain Disability Index* (PDI) is a 7-item questionnaire to investigate the magnitude of the self-reported pain related disability, independent from region of pain or pain-related diagnosis. The questionnaire is constructed on a 0-10 numeric rating scale in which 0 means no disability and 10 maximum disability. Total scores can range from 0 to 70, with higher scores reflecting higher interference of pain with daily activities. The PDI measures family / home responsibilities, recreation, social activity, occupation, sexual behaviour, self-care and life support activity [22, 23].

The *RAND-36 scale physical functioning* will be used to measure self-reported physical functioning independent of (pain) diagnosis [24]. The RAND-36 has been used widely across health conditions (www.rand.org, accessed august 2012). The physical functioning scale consists of 10 questions with 3 possible answers on a Likert scale: 'yes, strongly limited', 'yes, a bit limited', and 'no, not limited'. The total score can range from 0 to 100, with higher scores indicating better physical functioning. The validity and reliability of the Dutch version are good [25].

The *Work Reintegration Questionnaire* (WRQ) is an instrument for assessing the most important psychosocial factors in the delay of recovery and work resumption. The questionnaire consists of 78 items distributed among 8 scales; 'Distress', 'Illness behaviour', 'Job strain', 'Job dissatisfaction', 'Control', 'Avoidance', 'Perfectionism' and 'Stressful home situation.The *Work Reintegration Questionnaire* (WRQ) measures the following dimensions: distress, interference, work stress, work satisfaction, insecurity / avoidance, perfectionism / persistence, home situation [26]. The questionnaire was developed in Dutch (VAR: vragenlijst arbeidsreintegratie). A validated translation in English is currently in development (personal communication with author).

Pain intensity and fatigue can be assessed using an 11-point NRS (*NRS-pain* and *NRS-fatigue*), ranging from 0 (no pain / fatigue) to 10 (worst possible pain / fatigue), requiring patients to rate their current and average intensity of the last seven days [9].

Exercise tolerance functions will be assessed with standardized *lifting capacity tests* from the Workwell Functional Capacity Evaluation (FCE): lifting low and overhead lifting. Procedures are described in detail elsewhere [27]. These tests were found to be predictive of functional capacity performance in general in patients with back pain and neck / upper extremity pain [28]. A standardized

submaximal *Astrand bicycle test* [29, 30] or *Bruce treadmill test* [31] will be used to assess exercise tolerance functions as well as energy and drive functions.

The *Trimbos iMTA questionnaire for measuring Costs of Psychiatric Illnesses* (TiC-P), module 1, will be used to assess health care utilization. The questionnaire has a recall period of 4 weeks. Visits and consultations of the following health care providers were measured: general practitioner, physiotherapist, manual therapist, exercise therapist, occupational therapist, psychologist, insurance physician, medical specialists in hospitals, hospitalization

(number of days), occupational physician, social worker, and dietician. Further items were alternative care, home care, medication use, and job related care like job coaches, ergonomic changes at the work site and re-integration specialists [32]. Slight adaptations in the context and scope of health care practitioners were made to better fit TiC-P to the target population (i.e. from psychiatry to pain and work).

Global perceived effect (GPE) can be measured with a 7-point Likert scale ranging from 1 to 7 (1; 'extremely worsened', 2; 'much worsened', 3; 'little worsened', 4; 'unchanged', 5; 'little improved', 6; 'much improved', 7; 'completely improved'). Two GPE questions are proposed: how much did your treatment change your pain compared to pre-treatment level, and how much did your treatment change your work status compared to pre-treatment level?

Treatment records will be used to assess diagnosis, adherence to the treatment program, adverse effects that has not lead to discontinuation of the program, and reasons for premature withdrawal.

Table 3. Quality appraisal of VR-Pain core domains classified per ICF category

VR-Pain core domains	ICF code	Name of instrument or scale	Reliability	Construct validity	Responsive-ness	Formally validated translation	Utility (min)	References
Quality of Life[a]								
· Quality of life	x	EQ-5D	Yes	Yes	Yes	Yes	2	[33]
Activities / Participation								
· Remunerative employment	d850	WAI q3	Yes	Yes	AoE	Yes	<1	[20]
		PRODISQ	Yes	Yes	AoE	Yes	5	[21]
· Role functioning (i.e. work and educational activities)	d850	PDI q4	Yes	Yes	Yes	Yes	<1	[23,34]
· Physical functioning	d899	PDI total RAND-36 physical functioning	Yes	Yes	Yes	Yes	2	[23,24,34] [25]
· Complex interpersonal interactions	d720	WRQ satisfaction	Yes	Yes	NA	Yes	1	[26]
· Handling stress and other psychological demands	d240	WRQ work stress	Yes	Yes	NA	Yes	1	[26]
Body Functions								
· Pain	b280	Diagnosis	-	-	-	-	-	-
· Pain intensity	b280	NRS pain	Yes	Yes	Yes	Yes	<1	[35,36]
· Energy and drive functions	b130	NRS fatigue	Yes	Yes	Yes	No	<1	[40]
· Exercise tolerance functions	b455	Lifting test	Yes	Yes	AoE	NA	10	[27,28,37,38]
		Astrand bicycle ergometry	Yes	Yes	AoE	NA	15	[29,30]

Table 3. (Continued)

VR-Pain core domains	ICF code	Name of instrument or scale	Reliability	Construct validity	Responsive-ness	Formally validated translation	Utility (min)	References
		Bruce treadmill ergometry	Yes	Yes	AoE	NA	10	[39]
· Emotional functioning	b152	WRQ distress	Yes	Yes	NA	Yes	1	[26]
Environmental Factors								
· Immediate family	e310	PDI q1	Yes	Yes	NA	Yes	<1	[23,34]
		WRQ home	Yes	Yes	NA	Yes	1	[26]
· People in positions of authority	e330	WRQ satisfaction	Yes	Yes	NA	Yes	1	[26]
· Health care utilization	e580	TiC-P	Yes	Yes	NA	Yes	8	[32]
Personal factors								
· Coping	IM	WRQ avoidance	Yes	Yes	NA	Yes	1	[26]
		WRQ persistence	Yes	Yes	NA	Yes	1	[26]
Evaluation								
· Participant ratings of global improvement	IM	GPE	Yes	Yes	Yes	Yes	<1	[23]
· Adherence to the intervention and reasons for premature withdrawal; diagnosis;	IM	Medical records	AoE	AoE	AoE	NA	<1	-
· Clinician or surrogate ratings of global improvement	IM	GPE	AoE	AoE	AoE	NA	<1	-

[a]Not in IMMPACT or VR-Pain Core Sets; x=ICF code not available; IM= IMMPACT; EQ-5D= EuroQol-5D; WAI= work ability index; PRODISQ= PROductivity and DISease Questionnaire; PDI= pain disability index; WRQ= work reintegration questionnaire; GPE= global perceived effect; NA= not applicable; AoE= absence of evidence; TiC-P= Trimbos iMTA questionnaire for measuring Costs of Psychiatric Illnesses

Discussion

Sub-acute or chronic musculoskeletal pain can lead to a decrease in work participation up to the point where VR becomes essential. It is important to understand and address musculoskeletal pain in the context of VR because if we can mitigate the burden of work disability, we can facilitate early and sustained return to work. To do so, in this study, we attempted to blend two sources of domains around pain and VR, listed the instruments by which we can operationalize the domains, and developed a VR-Pain core set that may benefit clinical and research application in the VR-pain field in the Netherlands and potentially in other countries as well. To this end, the final VR-Pain core set consisted of 12 instruments that covered 18 domains.

As presented in Tables 1 and 2, domains and items of the underlying sets were removed by the authors because they were not deemed relevant for the (vast majority of) the target population. While this was not or incidentally challenged by members of the expert groups, this still needs elucidation. Acquiring skills (ICF code d155 *Acquiring skills*) for example, was excluded, because this it is not primarily affected (or core) in people with pain, and therefore not a goal in VR. *Acquiring skills* was deemed very relevant for the brief ICF Core Set for VR, because VR in its broadest form is provided to workers with a range of disabilities reaching far beyond pain, including workers with neurological and intellectual problems. Because of the specific setting for which the VR-Pain core set was developed, other items that were initially included to accommodate the wide application of both ICF and IMMPACT sets did not make to the core and final VR-pain set like unpaid work, acquiring, keeping, and terminating work, end-of-life issues, and higher level cognitive functions. To exclude the latter, however, may be subject to debate, because workers with pain often report challenges with concentration and memory. These concomitant complaints are regarded as related to pain and fatigue (which are already included in the final set), and perhaps symptoms related to central sensitization. Non-specific pain does not directly affect the brain and higher neurological functions as captured in ICF code b164 *Higher level cognitive functions*. For similar reasons biological markers were also excluded. Finally, while environmental issues such as insurance and social security systems are considered relevant [4] and vary across jurisdictions, they do not vary across the workers in the specific setting for which this VR-Pain core set was developed for. For generalizations beyond the Dutch borders, we advise researchers to describe the issues in future reports within the context or controlling for insurance and social security systems.

Most instruments proposed to make up the VR-Pain core set all comply with basic clinimetric properties as presented in Table 3. Because the properties included validated Dutch language versions of questionnaires and feasibility, the set of instruments proposed is likely to differ from core sets for different countries and languages. Additionally, we have attempted to choose instruments that could be used for clinical as well as for research purposes, including economic evaluations. Even though EQ-5D was not recommended by either one of the underlying core sets (ICF Core Set or IMMPACT), it was added because this instrument can be used for economic evaluations. Additionally, it captures an important secondary aim of vocational rehabilitation, which is to contribute to increase quality of life.

New core sets that apply to specific groups may thus be developed based on existing core sets. The exercise of developing a new core set based on two established ones has not been presented previously. Thus, the methodology described in this paper is new. We have aimed to describe this methodology transparently, to enable readers to either replicate these steps when developing or validating other core sets specific to their setting, or to use it as a basis for further development of this methodology. By asking Dutch experts in the pain rehabilitation field, prospective users and international VR experts, we aimed to test the content validity of the newly developed core set. However, this paper may also be regarded as external validation of the underlying core sets. In choosing the instruments, we aimed to combine sound psychometric properties with the options for future cost-effectiveness studies or intervention trials. This will enable future users to study clinical and economic outcomes in the (Dutch) usual care setting, which should make a significant contribution to the field of VR and pain.

While the response rates of the Dutch pain rehabilitation experts and prospective users was high, response of the international experts was low. Some international experts responded that this exercise was specifically not in their sub-field of expertise (e.g. cognitive vocational rehabilitation), the majority of this group did not respond at all, which may be attributed to lack of time availability or were unable to follow up on the electronic invitation and reminder. The relevance of this non-response is unknown. Because based on the responses only small changes were made to the final core set, and no differences in response patterns between expert groups were observed, we assume that the relevance of the non-response to be limited. Patients were not invited to participate in this specific exercise, because patient involvement was already incorporated in the development of the two underlying core sets. Both the ICF Core Set for

Vocational Rehabilitation and the IMMPACT recommendations are in principle experts-based. Even though the VR-Pain core set was agreed upon by most experts ('externally validated'), it is in its essence also an experts-based core set. Future use will discover whether the set of instruments is deemed too extensive for routine clinical use, and where and why this core set should be adapted to new developments in the VR-Pain field.

In line with the recommendations underlying the ICF VR expert group [12], the lack of classification of the personal factors in the ICF which can play a crucial influence on work functioning, will need careful consideration in the future. Although some performance-based instruments are included in the final VR-Pain core set, the majority of the instruments are self-report based. Apart from its strengths (outcomes are judged by the patients, not by or interpreted by others), this may also introduce a risk of bias, particularly in the estimation of absenteeism and presenteeism. Additionally, while clinimetric properties of the individual instruments in the VR-Pain core set were checked, they were not systematically reviewed.

The clinical relevance of using this VR-Pain core set is that it will provide a firm base for routine clinical use and evaluation of services in vocational rehabilitation settings with pain-related cases. Clinicians can, based on their clinical expertise or professional guidelines, add diagnostic instruments to this core set as needed. Moreover, the VR-Pain core set should not replace clinical expertise, but rather should complement it. The methodology described in this paper may be generalizable to develop other setting-specific core sets or a combination thereof. Additionally, most of the instruments in the VR-Pain core set are used internationally, which will address generalizability and comparability. Costs calculations underlying the EQ-5D, PRODISQ and TiC-P questionnaires, however, are based on Dutch guidelines which are expected to be different from other countries. While the VR-Pain core set is developed for the Netherlands, the burden of pain and work disability in the Netherlands is similarly high as in other industrialized countries [1]. Therefore, it is recommended that similar core sets are to be developed and tested for different countries. To enable generalization across countries, facilitate common language and stimulate future developments, we recommend that whenever possible, the same instruments are used.

Conclusion

A VR-Pain core set with content that covers relevant domains for pain and VR and with validated corresponding instruments that measure these domains has been developed. The VR-Pain core set may be used for clinical purposes, and (cost)effectiveness research in the field of vocational rehabilitation and pain. Caution is warranted for direct use outside of The Netherlands, because differences in cultural and service and political systems exist, hence the basis for costs calculations may be different. Additionally, for use and generalization beyond that of The Netherlands, it is recommended that environmental factors (ICF e580 and e590, Table 1) be considered and examined.

Acknowledgements

The authors wish to acknowledge the input of all responders. The following responders have granted permission to acknowledge them:

Pain rehabilitation experts: Karlein Schreurs, Rita Schiphorst Preuper, Albere Köke, Sylvia Remerie.

Prospective users: Peter van der Wurff, Martin Smeulers, Loes Swaan, Levijn Romp, Claire Tilmans, Sieger de Vries, Marleen ter Haar.

VR experts: Alex Burdorf, Debra Homa, Sven-Uno Marnetoft.

References

1. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain 2006 May;10(4):287-333.
2. Reid KJ, Harker J, Bala MM, Truyers C, Kellen E, Bekkering GE, et al. Epidemiology of chronic non-cancer pain in Europe: narrative review of prevalence, pain treatments and pain impact. Curr Med Res Opin 2011 Feb;27(2):449-462.
3. Bekkering GE, Bala MM, Reid K, Kellen E, Harker J, Riemsma R, et al. Epidemiology of chronic pain and its treatment in The Netherlands. Neth J Med 2011 Mar;69(3):141-153.
4. Anema JR, Schellart AJ, Cassidy JD, Loisel P, Veerman TJ, van der Beek AJ. 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 Dec;19(4):419-426.
5. Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. Spine J 2008 Jan-Feb;8(1):8-20.
6. Hoy D, March L, Brooks P, Woolf A, Blyth F, Vos T, et al. Measuring the global burden of low back pain. Best Pract Res Clin Rheumatol 2010 Apr;24(2):155-165.
7. Norlund A, Ropponen A, Alexanderson K. Multidisciplinary interventions: review of studies of return to work after rehabilitation for low back pain. J Rehabil Med 2009 Feb;41(3):115-121.
8. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, et al. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. Eur Spine J 2006 Mar;15 Suppl 2:S192-300.
9. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. Pain 1986 Oct;27(1):117-126.
10. Jensen IB, Busch H, Bodin L, Hagberg J, Nygren A, Bergstrom G. Cost effectiveness of two rehabilitation programmes for neck and back pain patients: A seven year follow-up. Pain 2009 Apr;142(3):202-208.
11. Escorpizo R, Reneman MF, Ekholm J, Fritz J, Krupa T, Marnetoft SU, et al. A conceptual definition of vocational rehabilitation based on the ICF: building a shared global model. J Occup Rehabil 2011 Jun;21(2):126-133.
12. Finger ME, Escorpizo R, Glassel A, Gmunder HP, Luckenkemper M, Chan C, et al. ICF Core Set for vocational rehabilitation: results of an international consensus conference. Disabil Rehabil 2012;34(5):429-438.
13. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain 2003 Dec;106(3):337-345.
14. Turk DC, Dworkin RH, Revicki D, Harding G, Burke LB, Cella D, et al. Identifying important outcome domains for chronic pain clinical trials: an IMMPACT survey of people with pain. Pain 2008 Jul 15;137(2):276-285.
15. Wind H, Gouttebarge V, Kuijer PP, Frings-Dresen MH. Assessment of functional capacity of the musculoskeletal system in the context of work, daily living, and sport: a systematic review. J Occup Rehabil 2005 Jun;15(2):253-272.

16. Lamers LM, McDonnell J, Stalmeier PF, Krabbe PF, Busschbach JJ. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ 2006 Oct;15(10):1121-1132.

17. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. Kwaliteit van leven meten in economische evaluaties: het Nederlands EQ-5D-tarief (Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff). Ned Tijdschr Geneeskd 2005, 149(28):1574-1578.

18. Bieleman HJ, Reneman MF, van Ittersum ,M.W., van der Schans ,C.P., Groothoff JW, Oosterveld FGJ. Self-reported functional status as predictor of observed functional capacity in subjects with early osteoarthritis of the hip and knee: a diagnostic study in the CHECK cohort. J Occup Rehabil 2009 12;19(4):345-353.

19. Lamers LM, Meerding WJ, Severens JL, Brouwer WB. The relationship between productivity and health-related quality of life: an empirical exploration in persons with low back pain. Qual Life Res 2005 Apr;14(3):805-813.

20. 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 Sep;36(5):404-412.

21. Koopmanschap M, Meerding WJ, Evers S, Severens J, Burdorf A, Brouwer W. Handleiding voor het gebruik van PRODISQ versie 2.1 (Manual for the use of the PRODISQ version 2.1). Rotterdam: 2004.

22. Tait RC, Chibnall JT. Factor structure of the pain disability index in workers compensation claimants with low back injuries. Arch Phys Med Rehabil 2005 Jun;86(6):1141-1146.

23. Soer R, Reneman MF, Vroomen PC, Stegeman P, Coppes MH. Responsiveness and minimal clinically important change of the Pain Disability Index in patients with chronic back pain. Spine (Phila Pa 1976) 2012 Apr 15;37(8):711-715.

24. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. Health Econ 1993 Oct;2(3):217-227.

25. Zee Kvd, Sanderman R. Het meten van de algemene gezondheidstoestand met de RAND-36, een handleiding (Measuring the general health status with the RAND-36, a manual). 2nd ed. Groningen, The Netherlands: RuG, research institute SHARE; 2012.

26. Vendrig A. De vragenlijst arbeidsreïntegratie (The Work Reintegration Questionnaire). Diagnostiek-Wijzer 2005;8:27-39.

27. Soer R, van der Schans C,P., Geertzen JH, Groothoff JW, Brouwer S, Dijkstra PU, et al. Normative values for a functional capacity evaluation. Arch Phys Med Rehabil 2009 10;90(10):1785-1794.

28. Gross DP, Battie MC. Construct validity of a kinesiophysical functional capacity evaluation administered within a worker's compensation environment. J Occup Rehabil 2003 Dec;13(4):287-295.

29. Keller A, Hellesnes J, Brox JI. Reliability of the isokinetic trunk extensor test, Biering-Sorensen test, and Astrand bicycle test: assessment of intraclass correlation coefficient and critical difference in patients with chronic low back pain and healthy individuals. Spine (Phila Pa 1976) 2001 Apr 1;26(7):771-777.

30. Hodselmans AP, Dijkstra PU, Geertzen JH, van der Schans CP. Exercise capacity in non-specific chronic low back pain patients: a lean body mass-based Astrand bicycle test; reliability, validity and feasibility. J Occup Rehabil 2008 Sep;18(3):282-289.

31. Bruce RA, Kusumi F, Hosmer D. Maximal oxygen intake and nomographic assessment of functional aerobic impairment in cardiovascular disease. Am Heart J 1973 Apr;85(4):546-562.

32. Bouwmans C, De Jong K, Timman R, Hakkaart-Van Roijen L. Feasibility and Validity of the TiC-P. Value in Health 2011;14(7):A426.

33. Dyer MT, Goldsmith KA, Sharples LS, Buxton MJ. A review of health utilities using the EQ-5D in studies of cardiovascular disease. Health Qual Life Outcomes 2010 Jan 28;8:13.

34. Gronblad M, Hupli M, Wennerstrand P, Jarvinen E, Lukinmaa A, Kouri JP, et al. Intercorrelation and test-retest reliability of the Pain Disability Index (PDI) and the Oswestry Disability Questionnaire (ODQ) and their correlation with pain intensity in low back pain patients. Clin J Pain 1993 Sep;9(3):189-195.

35. Cleland JA, Childs JD, Whitman JM. Psychometric properties of the Neck Disability Index and Numeric Pain Rating Scale in patients with mechanical neck pain. Arch Phys Med Rehabil 2008 Jan;89(1):69-74.

36. Young IA, Cleland JA, Michener LA, Brown C. Reliability, construct validity, and responsiveness of the neck disability index, patient-specific functional scale, and numeric pain rating scale in patients with cervical radiculopathy. Am J Phys Med Rehabil 2010 Oct;89(10):831-839.

37. Reneman MF, Dijkstra PU, Westmaas M, Goeken LN. Test-retest reliability of lifting and carrying in a 2-day functional capacity evaluation. J Occup Rehabil 2002 Dec;12(4):269-275.

38. van Abbema R, Lakke SE, Reneman MF, van der Schans CP, van Haastert CJ, Geertzen JH, et al. Factors associated with functional capacity test results in patients with non-specific chronic low back pain: a systematic review. J Occup Rehabil 2011 Dec;21(4):455-473.

39. Nordrehaug JE, Danielsen R, Stangeland L, Rosland GA, Vik-Mo H. Respiratory gas exchange during treadmill exercise testing: reproducibility and comparison of different exercise protocols. Scand J Clin Lab Invest 1991 Nov;51(7):655-658.

40. Minnock P, Kirwan J, Bresnihan B. Fatigue is a reliable, sensitive and unique outcome measure in rheumatoid arthritis. Rheumatology 2009;48:1533–1536.

CHAPTER 3

Test-retest reliability, agreement and responsiveness of productivity loss (iPCQ-VR) and healthcare utilization (TiCP-VR) questionnaires for sick workers with chronic musculoskeletal pain

T. Beemster
J. van Velzen
C. van Bennekom
M. Reneman
M. Frings-Dresen

J Occup Rehabil 2019 Mar;29(1):91-103

Abstract

Purpose

The purpose of this study was to assess test-retest reliability, agreement, and responsiveness of questionnaires on productivity loss (iPCQ-VR) and healthcare utilization (TiCP-VR) for sick-listed workers with chronic musculoskeletal pain who were referred to vocational rehabilitation.

Methods

Test-retest reliability and agreement was assessed with a two-week interval. Responsiveness was assessed at discharge after a 15-week vocational rehabilitation (VR) program. Data was obtained from six Dutch VR centers. Test-retest reliability was determined with intraclass correlation coefficient (ICC) and Cohen's kappa. Agreement was determined by Standard Error of Measurement (SEM), smallest detectable changes (on group and individual level), and percentage observed, positive and negative agreement. Responsiveness was determined with area under the curve (AUC) obtained from receiver operating characteristic (ROC).

Results

A sample of 52 participants on test-retest reliability and agreement, and a sample of 223 on responsiveness were included in the analysis. Productivity loss (iPCQ-VR): ICCs ranged from 0.52 to 0.90, kappa ranged from 0.42 to 0.96, and AUC ranged from 0.55 to 0.86. Healthcare utilization (TiCP-VR): ICC was 0.81, and kappa values of the single healthcare utilization items ranged from 0.11 to 1.00.

Conclusions

The iPCQ-VR showed good measurement properties on working status, number of hours working per week and long-term sick leave, and low measurement properties on short-term sick leave and presenteeism. The TiCP-VR showed adequate reliability on all healthcare utilization items together and medication use, but showed low measurement properties on the single healthcare utilization items.

Keywords

Productivity loss, Healthcare utilization, Vocational Rehabilitation, Cost-effectiveness, Measurement properties.

Introduction

Chronic musculoskeletal pain (CMP) is a common condition that results in major disability and substantial healthcare costs [1, 2]. CMP has a negative impact on performing work, resulting in productivity loss from work; reflected by absenteeism (sick off work) or presenteeism (productivity loss while at work) [3]. Productivity loss is labeled in cost-effectiveness studies as indirect healthcare costs [4]. Direct health costs are intervention costs, traveling costs and healthcare utilization costs. Vocation rehabilitation (VR) showed (cost-) effective in improving absenteeism and presenteeism and the reduction of healthcare utilization [5-7].

For clinical practice and research purposes, data about the (cost-)effectiveness of VR interventions are often collected with patient-reported outcome measures (PROMS). PROMS are standardized, validated questionnaires that are completed by patients to measure their perceptions of their functional status and wellbeing [8]. To give reliable statements on the (cost-)effectiveness of VR, PROMS on productivity loss and healthcare utilization must show adequate measurement properties [3, 8].

However, currently there are no gold standards available for the assessment of productivity loss [9-12]. Evidence on retest reliability and responsiveness on PROMS on absenteeism is scarce [13] and shows mixed results [11]. Research on retest reliability of five presenteeism questionnaires showed moderate to sufficient retest reliability in a sample with rheumatic diseases (ICCs 0.59-0.78) [10], and low to moderate responsiveness in a sample with rheumatoid arthritis or osteoarthritis [14]. However, some issues with presenteeism questionnaires are prominent; they have different recall periods, different outcome scales (0-10 or 1-7), are developed for different populations (general or sickness-specific, for example rheumatic diseases), and they measure different concepts of presenteeism, for example productivity, performance or ability [10]. As a consequence, the correlation between global measures of presenteeism is low, which complicates comparison [10].

Two Dutch questionnaires on the assessment of productivity loss and healthcare utilization have recently been developed. These questionnaires are recommended by the Dutch guideline for health economic evaluations [4]. The questionnaire on the measurement of productivity loss is called the iMTA Productivity Cost Questionnaire (iPCQ) [11, 15-17] and the questionnaire on the assessment of

healthcare utilization is called the Trimbos iMTA questionnaire for measuring Costs of Psychiatric Illnesses (TiC-P, part I) [18]. In addition, the TiC-P consists of two parts, a healthcare usage part (part I) and a productivity loss part (part II). Part II has been further developed for the general population and resulted in the iPCQ. In a sample with mental problems, the TiC-P (parts I and II) showed sufficient feasibility and construct validity, and low to sufficient retest reliability [18]. In another study, the feasibility and face validity of the iPCQ was confirmed [15].

However, the iPCQ and TiC-P questionnaires are not fully applicable for sick workers with CMP who are referred to VR. For example, a large portion of sick workers referred to VR are on part-time sick leave and thus part-time at work. The iPCQ, however, does not measure part-time work/sick leave. Furthermore, the TiC-P questionnaire contains many items about mental healthcare but, for example, no items about workplace adaptations or visits of reintegration specialists. Therefore, we modified the iPCQ and TiCP questionnaires to enhance feasibility and usefulness. We called these modified versions the TiCP-VR and the iPCQ-VR. The aim of this study is to assess the test-retest reliability, agreement and responsiveness of the iPCQ-VR and TiCP-VR in workers with chronic musculoskeletal pain and referred to VR in the Netherlands.

Methods

The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist was applied in the design of the study [19].

Procedures

For this study we used two study samples. The first study sample was used to perform the retest reliability and agreement analysis, the second study sample was used to perform the responsiveness analysis. Participants of the first sample were recruited from six VR centers in the Netherlands (Rijndam, MRC Doorn, Klimmendaal, Trappenberg, UMCG CvR and Heliomare). At baseline (T0), patients completed the iPCQ-VR, TiCP-VR and other web-based questionnaires at home as part of care as usual [20]. After a multidisciplinary screening, eligible patients were informed about the study by a member of the multidisciplinary screening team and written information describing the study was provided. Two weeks after T0, respondents received the iPCQ-VR and TiCP-VR for the second time (T1). If T0 was more than two weeks before granted informed consent, the T0 and T1 questionnaires were sent with two weeks in between. If participants did not complete the T0 or T1 questionnaires within a week, they received a reminder email. If the questionnaires were not completed after this reminder, participants were phoned by the first author TB. Data of study sample 2 was derived from routinely collected data from six Dutch rehabilitation centers (Heliomare, Roessingh, Adelante, Libra, Klimmendaal, Trappenberg), all offering a multidisciplinary VR program (15-week duration) for workers with chronic musculoskeletal pain. We used baseline (T0) and discharge data (T2). The T2 questionnaires were automatically sent 14 weeks after the start of the VR program. Figure 1 shows the measurement points of samples 1 and 2.

Participants

The inclusion criteria were: 1) being of working age (18 to 65 years); 2) suffering from subacute (6 to 12 weeks) or chronic (>12 weeks) nonspecific musculoskeletal pain such as back, neck, shoulder, widespread pain, Whiplash Associated Disorder (WAD I or II), or fibromyalgia; 3) having paid work (employed or self-employed) for at least 12 hours per week; 4) having sick leave (part-time or full-time); 5) being able to complete questionnaires in Dutch; 6) having an email address; and 7) having granted informed consent. The exclusion criterion was having comorbidities that were the primary reason for sick leave, such

as acute or specific medical problems, clinical depression or burnout, severe asthmatic symptoms, diagnosed chronic fatigue, and neuropathy. The Medical Ethical Committee of the Academic Medical Center, Amsterdam, the Netherlands, authorized this study and decided that a full application was not required. Participation in the study was voluntary, all participants provided informed consent and answers were processed anonymously.

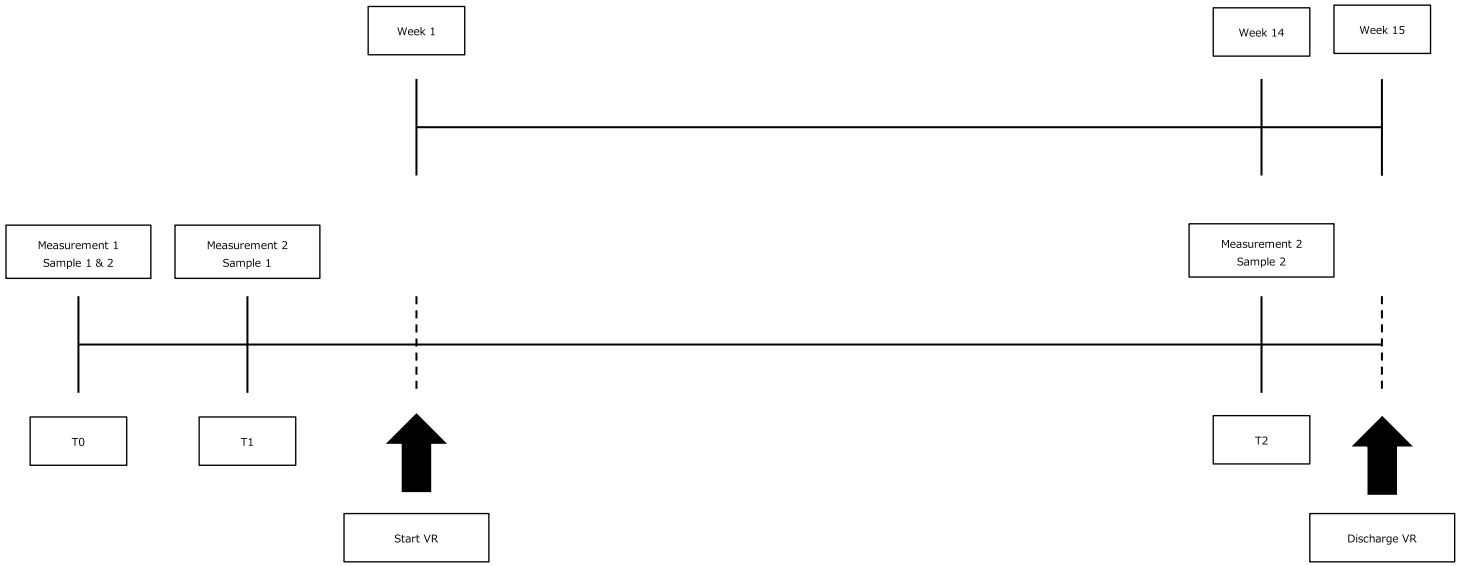
Measurements

Patient characteristics
Several demographic and clinical variables were assessed at baseline: age, gender, education, pain features (location, duration and intensity), work features (status, contract), and level of disability.

iPCQ-VR
The iPCQ-VR is a modified version of the iPCQ [11, 15, 17, 18], and is used by six VR centers in the Netherlands. The iPCQ-VR adopted the absenteeism and presenteeism modules of the original iPCQ [17], and two extra modules were added: working status and pain-specific sick leave. We pilot-tested preliminary versions within our research team and four patients pilot-tested the pre-final version of the questionnaire. All items of the iPCQ-VR and the corresponding rating scales are shown in Appendix 1.

TiCP-VR
The original TiC-P assesses the visits and consultation of several healthcare providers, and medication use [18]. The utilization of each healthcare provider is assessed with a yes/no item and if patients answer 'yes', the number of visits/consultations is assessed. A recall period of 4 weeks is used in the original questionnaire, which we adopted in the TiCP-VR version. In the TiCP-VR version, we removed five items that were specific to psychiatric patients, but not for our population. Furthermore, we added pain-specific items to allow differentiation between pain-related and other healthcare utilization. Finally, we removed non pain-related medication use. This was due to feasibility reasons and it was expected that medication use other than pain-related was marginal when translated to costs. Also, it was expected that this adaptation would prevent missing data on medication use, as this was prominent in the original TiC-P validation study [18]. We pilot-tested preliminary versions within our research team and four patients pilot-tested the pre-final version of the questionnaire. All items of the TiCP-VR and the corresponding rating scales are shown in Appendix 2.

Figure 1. Measurement points of this study



VR vocational rehabilitation
Sample 1: assessment of test-retest reliability and agreement
Sample 2: assessment of responsiveness

Global perceived effect

One global perceived effect (GPE) item ('How much did the vocational rehabilitation program change your work functioning compared to pre-treatment level?') was assessed at T2 and was used as the external criterion (anchor) in the responsiveness analysis in this study. GPE was measured with a 7-point Likert scale ranging from 1 to 7 (1; 'extremely worsened', 2; 'much worsened', 3; 'little worsened', 4; 'unchanged', 5; 'little improved', 6; 'much improved', 7; 'completely improved').

Statistical Analysis

Reliability

Test-retest reliability of the *continuous items* of the iPCQ-VR were performed with intraclass correlation coefficient (ICC random, single, and on absolute agreement) [21]. To allow comparison with other studies, in particular the original iPCQ study by Bouwmans et al. [18], we performed sensitivity analyses with ICC random, average, and on absolute agreement. One overall ICC of all healthcare visits/consultations of TiCP-VR together was calculated because the single continuous items were expected to be underpowered [18].We considered an ICC of >0.70 sufficient for use at group level and an ICC of >0.90 sufficient for use at individual level [22].

Reliability of *dichotomous items* of iPCQ-VR and TiCP-VR were studied using Cohen's kappa analyses $\left[k = \frac{Po-Pc}{1-Pc}\right]$ where *Po* is the proportion of observed agreements and *Pc* is the proportion of agreements expected by chance [23]. The range of possible values of kappa is from –1 to 1 [23]. We interpreted kappa values as follows: slight (0.00-0.20), fair (0.21-0.40), moderate (0.41-0.60), substantial (0.61-0.80) and almost perfect (0.81-1.00) [24]. The pain-specific items of the TiCP-VR were expected to be underpowered and were blended to one 2 x 2 contingency table.

Reliability of *categorical variables* was performed with linear weighted kappa coefficients [25, 26].

Agreement

Agreement of *continuous variables* was analyzed by the Standard Error of Measurement $\left[SEM = SD\sqrt{1} - ICC\right]$, where SD is the SD of the scores from all participants, which were determined from an ANOVA analysis with the formula $\left[\sqrt{SStotal\,/(n-1)}\right]$, and ICC is the retest reliability coefficient [21]. The

SEM was converted into the smallest detectable changes on individual level $\left[SDC_{individual} = 1.96 * \sqrt{2} * SEM\right]$. This number reflects the smallest within-person change in a score that can be considered to be a real change above any measurement error within one individual. The SDC individual was converted into the SDC for a group (SDC group) by dividing the SDC individual by $\sqrt{n}$. We proposed a positive rating for agreement if the absolute measurement error (SDC individual for change within individuals and SDC group for change between groups) is smaller than the minimal important change (MIC, see responsiveness) [27, 28].

Agreement of *dichotomous variables* was analyzed by the percentage observed agreement $\left[P_o = \frac{(a+d)}{n}\right]$, the percentage positive agreement $\left[PA = \frac{2a}{2a+b+c}\right]$, and the percentage negative agreement $\left[NA = \frac{2d}{2d+b+c}\right]$ [29]. PA is known as the specific agreement on a positive rating and NA is known as the specific agreement on a negative rating [29]. All 2 x 2 contingency tables will be provided in Appendices 3 and 4. *Categorical variables* were analyzed by the percentage observed agreement.

Responsiveness

Responsiveness in this study was defined as the ability of the iPCQ-VR to detect clinically relevant changes over time [27]. We assessed the responsiveness on four continuous items: the number of sick leave days in the preceding 4 weeks (for participants with short-term sick leave at T0), the number of working hours per week (for participants with 100% sick leave at T0), the number of presenteeism days in the preceding 4 weeks and the presenteeism score (0-10) (for participants who scored 'yes' on presenteeism at T0). Various statistics were applied to calculate responsiveness [30]. Mean changes and 95% confidence intervals of mean changes were calculated. Sensitivity and specificity for change plotted by receiver operating characteristic (ROC) curve and area under the curves (AUCs) were calculated [31]. The AUC is the probability of correctly discriminating between improved and nonimproved patients. When the AUC was more than 0.70, responsiveness was considered sufficient [27]. MIC was measured by determining the optimal cut-off point (OCP). This is the point of the ROC curve where the sum of sensitivity and 1-specificity is maximal. Sensitivity and specificity of the OCP were computed. Sensitivity and specificity range from 0 to 1.00, where higher numbers reflect higher sensitivity or specificity. Because the objective of the responsiveness analysis was to differentiate between improved and unchanged samples of participants, the GPE score was dichotomized into a subgroup with GPE score "improved" (little improved,

much improved and completely improved) and a subgroup with the GPE score "unchanged". The GPE group "worsened" was not included in the analyses [30].

Stability
The ICC, kappa, and agreement analyses were performed on a stable sample that completed the questionnaire twice in similar conditions, with a two-week interval. To perform this, we added external anchor items at T1 (external anchor item: 'In relation to question x, did something change in the preceding two weeks, compared to the weeks before?'). To allow comparison with other studies, results of both stable and unstable (i.e. total sample) retest samples will be reported.

We applied an online calculation tool to calculate kappa and linear weighted kappa [32]. All other analyses were performed using SPSS 23 for Windows (SPSS Inc., Chicago, USA). The demographic data of the individuals were described by means and standard deviations (SD), or inter-quartile range in the case of no normal distribution. The assumption of normal data distribution was visually verified using histograms and QQ-plots.

Power
Fifty patients are needed to obtain a reasonable 2 x 2 contingency table to determine the kappa and to obtain a confidence interval ranging from 0.70-0.90 around an ICC of 0.80 [12, 24, 27]. Fifty to 99 patients are needed to obtain reasonable responsiveness scores [33].

Results

A total of 52 participants completed the retest questionnaires (response rate retest 71%). Reasons for non-response were technical problems (n=7), withdrawal consent (n=3), no telephone number (n=2), or unknown (n=9). The retest was submitted on average 19.6 days (SD 5.8) after submission of the initial questionnaires. A sample of 223 participants completed baseline and discharge responsiveness questionnaires. Response rates of this sample were unknown. The responsiveness questionnaires were submitted on average 14.5 weeks (SD 1.0) after T0. Table 1 shows the characteristics of both study samples.

Table 1. Characteristics of the study populations

	Reliability and agreement (n=52) Mean (SD) or %	Responsiveness (n=223) Mean (SD) or %
Age (years)	44.6 (10.3)	47.4 (10.9)
Gender (% female)	78.8	59.9
Education		
Low	17.0	15.3
Medium	55.3	55.4
High	27.7	24.3
Work (% yes)	96.2	96.0
Work situation		
Employed	90.2	87.5
Self-employed	2.0	2.7
Disability benefit	3.9	1.4
Housewife/houseman	2.0	1.4
Incapacitated	2.0	0.9
Contract (hours/week)	32.8 (7.0)	31.3 (8.8)
Contract (days/week)	4.3 (1.1)	4.2 (1.0)
Work status		
Working full-time	14.0	8.5
Working part-time	58.0	52.6
100% sick leave	28.0	39.0
Sick leave short (% yes)	50	62
Sick leave long (% yes)	36	49.8
Presenteeism (% yes)	74	63.8
Pain location		
Spine (% yes)	69.2	76.7
Lower extremities (% yes)	32.7	35.0
Upper extremities (% yes)	42.3	45.3
Headache/burnout	30.8	40.8
Number of pain locations (IQR)	2 (1-4)	2 (1-4)
Pain duration		
1-3 months	9.6	8.6
3-6 months	15.4	15.3
0.5-1 year	28.8	23.0
1-2 years	15.4	22.1
2-5 years	17.3	13.1
More than 5 years	13.5	18.0
Pain mean (0-10) [a]	6.0 (1.9)	5.5 (2.3)
Pain worse (0-10) [a]	7.4 (1.9)	6.9 (2.5)
Pain Disability Index (0-70) [b]	37 (12.1)	34.3 (11.6)

SD standard deviation, IQR interquartile range
[a] 0=no pain, 10=worst possible pain, [b] 0=no disability, 70=maximum disability

Reliability

The ICCs of the iPCQ-VR ranged from 0.52 to 0.90 (Table 2). Number of working hours per week scored 0.90, number of short-term sick leave days scored 0.54, presenteeism score scored 0.56, and number of presenteeism days scored 0.52. The ICC of total healthcare utilization was 0.81. Sensitivity analysis with average measures of ICC showed the following ICCs: number of working hours (0.95), presenteeism score (0.72), number of presenteeism days (0.68), number of sick leave days (0.70), and total healthcare utilization (0.89).

Cohen's kappa of the iPCQ-VR ranged from 0.42 to 0.96 (Table 2). In the total (both stable and unstable participants) sample, long-term pain-specific sick leave scored a kappa of 1.00 (Table 3). Cohen's kappa items of the healthcare utilization items of the TiCP-VR ranged from 0.11 to 1 (Table 4). Medication use showed substantial kappa (0.78) and total pain-specific healthcare utilization showed fair kappa (0.35). Table 5 shows kappa and agreement measures of the total sample on the TiCP-VR items. Appendix 3 (iPCQ-VR) and Appendix 4 (TiCP-VR) show all 2 x 2 contingency tables of both stable and unstable (total) samples.

Agreement

For the continuous items of the iPCQ-VR, the SEM, SDCind and SDCgrp were respectively 0.8, 2.3, 0.6 (number of working hours per week), 3.6, 10.1, 2.5 (number of sick leave days), 2,8, 7.9, 1.6 (number of presenteeism days), 0.7, 2.0, 0.4 (presenteeism score) (Table 6).

For the dichotomous items, observed agreement of the iPCQ-VR ranged from 72-98%, positive agreement ranged from 71-96% and negative agreement ranged from 62-91% (Table 2). Observed agreement (OA) of the healthcare items of the TiCP-VR ranged from 56-100%, positive agreement (PA) ranged from 48-100%, and negative agreement (NA) ranged from 39-100% (Table 4). Medication use scored OA: 89%, PA: 91%, NA: 87%. Pain-specific medication use (categorical item) scored OA: 59%. All pain-specific healthcare items together scored OA: 89%, PA: 94%, NA: 40%.

Responsiveness

The AUC, MIC, sensitivity and specificity of the iPCQ-VR are presented in Table 6 and the ROC curves are shown in figure 2. The AUCs ranged from 0.55-0.86.

The number of working hours per week showed adequate responsiveness for the participants who were on 100% sick leave at baseline (AUC 0.86, MIC = -1). Sick leave days in the preceding four weeks showed moderate responsiveness (AUC 0.66, MIC = 5.5). Presenteeism days in the preceding four weeks showed poor responsiveness (AUC 0.55, MIC = 4.5). Presenteeism score showed moderate responsiveness (AUC 0.60, MIC = -0.5 to -1.5). Table 7 shows the mean change scores of the iPCQ-VR.

Table 2. ICC, kappa and agreement of the iPCQ-VR in a stable group of participants

	Observed agreement (%)	Positive agreement (%)	Negative agreement (%)	Cohen's kappa (95% CI)	Yes (N)	ICC (95% CI)
Work status						
Working status	98	NA	NA	0.96 (0.88-1) [a]	41	-
Number of hours working per week [b]	-	-	-	-	16	0.90 (0.74-0.96)
Sick leave short (<4 weeks)						
Sick leave during last 4 weeks	72	71	73	0.45 (0.19-0.70)	16	-
Number of sick leave days	-	-	-	-	16	0.54 (0.05-0.81) [c]
Sick leave pain-specific	94	96	80	0.76 (0.32-1)	13	-
Sick leave long (>4 weeks)						
Sick leave longer than 4 weeks	88	83	91	0.74 (0.54-0.94)	15	-
Sick leave pain-specific	NA	NA	NA	NA	14	-
Presenteeism (<4 weeks)						
Presenteeism during last 4 weeks	74	80	62	0.42 (0.13-0.72)	22	-
Number of presenteeism days	-	-	-	-	23	0.52 (0.14-0.76)
Score of 0-10	-	-	-	-	23	0.56 (0.21-0.79)
Healthcare usage [d]	NA	NA	NA	NA	37	0.81 (0.66-0.89)

NA not applicable.
[a] Linear weighted kappa.
[b] In the case of part-time working (one of the three 'working status' options).
[c] One extreme outlier (value 49) was excluded from the analysis.
[d] All TiCP-VR healthcare usage items (01-14) together.

Table 3. ICC, kappa and agreement of the iPCQ-VR in total sample (stable and unstable participants)

	Observed agreement (%)	Positive agreement (%)	Negative agreement (%)	Cohen's kappa (95% CI)	Yes (N)	ICC (95% CI)
Work status						
Working status	86	NA	NA	0.80 (0.65-0.94) [a]	50	-
Working part-time: number of hours working per week [b]	-	-	-	-	24	0.76 (0.52-0.89)
Sick leave short (<4 weeks)						
Sick leave during last 4 weeks	72	71	73	0.44 (0.19-0.69)	17	-
Number of sick leave days	-	-	-	-	16	0.54 (0.05-0.81) [c]
Sick leave pain-specific	94	80	97	0.77 (0.34-1.00)	14	-
Sick leave long (>4 weeks)						
Sick leave longer than 4 weeks	88	83	91	0.74 (0.54-0.94)	15	-
Sick leave pain-specific	100	100	100	1.00 (1.00-1.00)	14	-
Presenteeism (<4 weeks)						
Presenteeism during last 4 weeks	76	83	60	0.43 (0.17-0.70)	29	-
Number of presenteeism days	-	-	-	-	29	0.51 (0.18-0.74)
Score of 0-10	-	-	-	-	29	0.54 (0.22-0.75)

NA not applicable.
[a] Linear weighted kappa.
[b] In the case of part-time working (one of the three 'working status' options).
[c] One extreme outlier (value 49) was excluded from the analysis.

Table 4. Kappa and agreement of the TiCP-VR in a stable group of participants

	N	Use of service (N)	No use of service (N)	Observed agreement (%)	Positive agreement (%)	Negative agreement (%)	Cohen's kappa (95% CI)
01 General Practitioner	46	9	17	56	57	63	0.13 (0-0.42)
02 PT, ET, OT, MT	42	16	16	76	86	76	0.55 (0.31-0.79)
03 Occupational physician	43	18	6	56	65	39	0.11 (0-0.41)
04 Reintegration specialist	44	6	32	86	67	91	0.59 (0.28-0.89)
05 Insurance physician	49	0	47	96	NA	98	NA
06 Workplace adaptations	48	6	35	85	63	91	0.54 (0.23-0.86)
07 Medical specialist	46	6	31	80	57	87	0.45 (0.12-0.77)
08 Stay in healthcare setting	50	0	50	100	NA	100	NA
09 Psychiatrist or psychologist	47	5	31	76	48	85	0.35 (0-0.68)
10 Social worker	49	0	48	98	NA	99	NA
11 Dietician	49	1	48	100	100	100	1
12 Home care	49	0	48	98	NA	99	NA
13 Alternative care	48	3	41	92	60	95	0.56 (0.14-0.97)
14 Medication use	38	21	13	89	91	87	0.78 (0.58-0.98)
Medication use pain-specific	27	14	2	59	NA	NA	0.31 (0-0.66) [a]
Healthcare usage pain-specific [b]	50	70	3	89	94	40	0.35 (0-0.75)

NA not applicable, PT, ET, OT, MT physical therapist, exercise therapist, occupational therapist, manual therapist, [a] Linear weighted kappa, [b] The sum of all pain-specific healthcare usage items.

Table 5. Kappa and agreement of the TiCP-VR in total sample (stable and unstable participants)

	N	Use of service (N)	No use of service (N)	Observed agreement (%)	Positive agreement (%)	Negative agreement (%)	Cohen's kappa (95% CI)
01 General Practitioner	50	12	18	60	55	64	0.21 (0-0.48)
02 PT, ET, OT, MT	50	23	16	78	81	74	0.57 (0.35-0.80)
03 Occupational physician	50	19	7	52	61	59	0.04 (0-0.32)
04 Reintegration specialist	49	9	33	86	72	90	0.63 (0.37-0.88)
05 Insurance physician	50	0	47	94	na	97	NA
06 Workplace adaptations	50	7	35	84	64	90	0.53 (0.24-0.83)
07 Medical specialist	50	8	33	82	44	88	0.52 (0.24-0.80)
08 Stay in health care setting	50	0	50	100	na	100	NA
09 Psychiatrist or psychologist	50	7	31	76	54	84	0.40 (0.10-0.69)
10 Social worker	49	0	48	98	na	99	NA
11 Dietician	49	1	48	100	100	100	1
12 Home care	49	0	48	98	na	99	NA
13 Alternative care	49	4	41	92	67	95	0.62 (0.27-0.98)
14 Medication use	48	30	13	90	92	84	0.76 (0.56-0.96)
Medication use pain specific	30	17	2	63	NA	NA	0.36 (0.05-0.68) [a]
Health care usage pain specific [b]	50	74	4	82	90	32	0.23 (0-0.56)

NA not applicable, PT, ET, OT, MT physical therapist, exercise therapist, occupational therapist, manual therapist, [a] Linear weighted kappa, [b] The sum of all pain-specific healthcare usage items.

Table 6. Responsiveness iPCQ-VR

Variable	N anchor	AUC (CI)	MIC	Sens	Spec	SEM [a]	SDCind [a]	SDCgrp [a]
Working hours [b]	49 positive 22 negative	0.86 (0.77-0.94)	-1	.796	.818	0.83	2.31	0.58
Sick leave days [c]	76 positive 31 negative	0.66 (0.55-0.78)	5.5	.671	.581	3.64	10.09	2.52
Presenteeism days [d]	84 positive 28 negative	0.55 (0.44-0.67)	4.5	.548	.571	2.84	7.87	1.64
Presenteeism score [d]	89 positive 29 negative	0.60 (0.48-0.72)	-0.5 to -1.5	.708 .584	.414 .655	0.71	1.98	0.41

[a] Available from the SDs and ICCs of sample I.
[b] Selection of participants who were on 100% sick leave at T0 and who reported stable or improved work functioning at T2.
[c] Selection of participants who were on short-term sick leave at T0 and who reported stable or improved work functioning at T2.
[d] Selection of participants who scored 'yes' on presenteeism at T0 and who reported stable or improved work functioning at T2.

Table 7. Mean change scores of iPCQ-VR

	Working hours	Sick leave days	Presenteeism days [a]	Presenteeism score [b]
Baseline T0				
Mean (SD)	0	13.1 (7.0)	11.8 (6.6)	5.3 (2.3)
Minimum-maximum	0-0	0-31	2-28	0-10
N	83	123	130	136
Discharge T2				
Mean (SD)	8.7 (10.7)	4.1 (7.4)	5.2 (6.0)	7.3 (2.5)
Minimum-maximum	0-40	0-31	0-20	0-10
N	81	127	129	131
Mean change (SD)	-8.7 (10.7)	9.2 (9.0)	6.4 (7.4)	-2.02 (3.2)
95% CI of mean change	-11.08 - -6.35	7.58 - 10.85	5.08 - 7.72	-2.58 - -1.46
N	81	119	124	131

[a] If presenteeism in the preceding 4 weeks was answered no on T2, presenteeism days was set at zero (0).
[b] If presenteeism in the preceding 4 weeks was answered no on T2, presenteeism score was set at ten (10).

Figure 2. ROC curves of the iPCQ-VR

2A. Number of working hours per week.

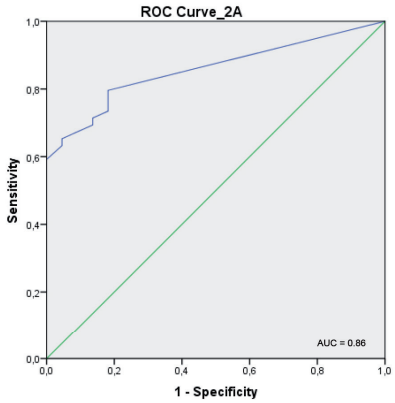


Fig. 2A ROC curve of participants who were on 100% sick leave at T0 and who reported stable or improved work functioning at T2 (n = 71)

2B. Number of sick leave days in preceding 4 weeks.

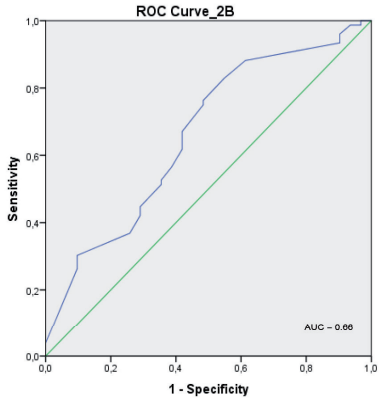


Fig. 2B ROC curve of participants who were on short-term sick leave at T0 and who reported stable or improved work functioning at T2 (n = 107)

2C. Number of presenteeism days in preceding 4 weeks.

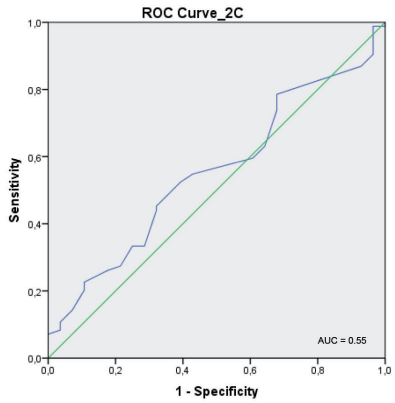


Fig. 2C ROC curve of participants who scored 'yes' on presenteeism at T0 and who reported stable or improved work functioning at T2 (n = 112)

2D. Presenteeism score (0-10) of preceding 4 weeks.

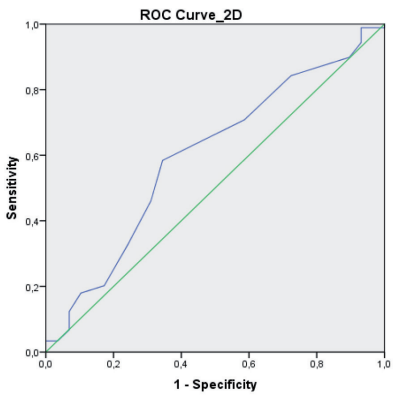


Fig. 2D ROC curve of participants who scored 'yes' on presenteeism at T0 and who reported stable or improved work functioning at T2 (n = 118)

ROC receiver operating characteristic, AUC area under the curve

Discussion

In this study, the retest reliability, agreement and responsiveness of two modified questionnaires on productivity loss (iPCQ-VR) and healthcare utilization (TiCP-VR) for workers on sick leave due to chronic musculoskeletal pain and referred to VR was assessed.

iPCQ-VR

The working status and number of working hours per week items scored high on retest reliability, agreement, and responsiveness. These items can be used at the group and individual levels as well as for evaluative purposes. Long-term sick leave scored sufficient retest reliability and agreement and can be used at group level. Short-term sick leave and presenteeism scored low retest reliability, agreement and responsiveness, and can therefore not be used at the group or individual level, or for evaluative purposes.

Reliability

Comparing the retest reliability of the absenteeism items of the current study with the original study [18] is complicated, because the original study used average measures ICC[2], which results in higher ICCs. In our opinion, single measures ICC is the appropriate ICC to answer the research question on retest reliability because in clinical practice patients complete the iPCQ-VR once per measurement point (i.e. at baseline, discharge, follow-up). Furthermore, the original study measured short-term sick leave with a recall period of two weeks, whereas we applied four weeks. Finally, the original study did not select a stable group of participants.

In a recent systematic review, the psychometric properties of eleven work productivity questionnaires were examined [11]. Data on the retest reliability of absenteeism was available for only four questionnaires. However, we cannot compare our results with these questionnaires for several reasons: no ICC or kappa performed [34-36], type of ICC unknown [37, 38], or a different recall period (3 months) and calculation of kappa (absenteeism 0 vs. >0 days) [39].

[2] The type of ICC is not clearly stated in the article. This information was known after e-mailing with the last author LHvR.

Despite the importance of absenteeism data as a return to work outcome and as a resource for economic evaluations, the evidence on the reliability of absenteeism measures is remarkably scarce. A possible explanation for this is that in several countries researchers can obtain sick leave data from social security databases [40], which is a feasible and reliable alternative [13]. However, such databases are not available for all countries, and another disadvantage is that the accuracy of sick leave data from electronic databases is low for short recall periods (i.e. ''acute'' sick leave) [12, 13, 41]. Because the reliability of short term sick leave was also low in the present study, this measure warrants improvement in future studies.

The ICCs ranging from 0.52 to 0.56 of the presenteeism items of the current study are somewhat lower compared with a review on the reliability of five at work productivity loss questionnaires in patients with rheumatic diseases, with single measures ICCs ranging from 0.59 to 0.78 (n=62-65) [10]. The higher ICCs of other studies can be explained by the low power (n=23) and longer recall period (four weeks) of the present study. A power of ≥50 and a recall period of 1 week is advocated [12].

Agreement

The observed agreement of the current study was somewhat lower compared with the original study (short-term sick leave: 72% vs. 87%, long-term sick leave: 88% vs. 93%, and presenteeism: 74% vs. 81%) [18]. This difference can be explained through a difference in power (n=50 vs. n=79). Unfortunately, the original study did not calculate the positive and/or negative agreement. There is one study known which also calculated observed agreement [39], but comparison with this study is not possible due to a different calculation of kappa (0 vs. >0 hours of absenteeism, presenteeism). As there are currently no cut-off scores available for the interpretation of positive and negative agreement, the information from the 2 x 2 contingency tables (Appendix 3) can be used by the reader to judge the uptake of a questionnaire or a particular item.

Responsiveness

The responsiveness analyses showed that a minimal important change of ≥1 working hours per week at discharge of VR can be used for evaluative purposes for patients who are on full sick leave at baseline. A minimal important change of 5.5 sick leave days per month can be considered for evaluative purposes for

patients who are on full sick leave at baseline. However, this warrants caution because the moderate AUC value of 0.66 is below the adequate level of 0.7.

The number of presenteeism days and the presenteeism score cannot be used for evaluative purposes because the AUCs were too low (0.55 and 0.60). One study assessed the responsiveness of five presenteeism scales (ranging from 0-10 or 1-7) [14]. In this study, ROCs and AUCs were assessed (and no MICs). The AUCs in this study ranged from 0.52-0.66, which is similar to that of the current study.

TiCP-VR

The sum of all healthcare visits of the TiCP-VR showed sufficient retest reliability and agreement, and can be used at group level. However, the single healthcare items of the TiCP-VR showed low kappa values and moderate agreement, which can be explained by uneven distributions of the 2 x 2 contingency tables (Appendix 4). This negatively affects the kappa and agreement values [23]. Furthermore, of four healthcare items (stay in a healthcare setting, social worker, insurance physician, home care) it was not possible to calculate kappa and agreement measures as none of the participants used these services. These items may be deleted to increase feasibility.

Medication use showed substantial retest reliability and adequate agreement. This item can be used at group level. In contrast, pain-specific medication use scored poor retest reliability and agreement, and this item cannot be used at group level and needs to be refined. Unfortunately, due to a technical error we were not able to assess the dosage, frequency and name of the consumed pain medications. The observed agreement of the current study is in line with the observed agreement from the original study [18]. Comparison on retest reliability (ICC values) with the original study is not possible as they used a different type of ICC.

Strengths and limitations

A strength of this study is that we included a sample of patients with chronic musculoskeletal pain who were referred to six VR centers in the Netherlands. This increases the clinical utility of this study. Second, we have extensively investigated both PROMS and we provided all 2 x 2 contingency tables (Appendices 3-4), as recommended [29].

Our results should be generalized cautiously as our study has some limitations that must be addressed. First, an inclusion criteria for this study was that participants should be on sick leave (part-time or full-time) at baseline. However, 14% of study sample one and 8.5% of study sample two were not at sick leave at baseline but full-time at work. This has resulted in lower samples for the performed analyses, which probably negatively affected the results on sick leave and presenteeism. Second, we applied anchor items at measurement 2 to detect stable and unstable (i.e. changed) samples of participants. For working status and the number of hours working per week, this resulted in better results on retest reliability in the stable group of participants. However, for the other items of the iPCQ-VR, such as short- and long-term sick leave and presenteeism, the results remained the same. Remarkably, the healthcare items of the TiCP-VR showed in general lower retest reliability (lower kappa values) in the stable sample compared with the unstable sample. Therefore, the anchor items applied in this study warrant refinement. Third, we assessed presenteeism with a time interval of two weeks. This is in line with similar studies [10]. Presenteeism may be unstable; it can fluctuate between days and weeks. Sim et al. [23] stated that for the time interval in retest reliability studies 'the stability of the attribute being rated is crucial to the period between repeated ratings'. We advise using a shorter time interval (for example two days) with control for stability to increase retest reliability in future studies.

The fourth and final limitation is the second measurement point in the responsiveness analysis (figure 1). Due to feasibility/technical reasons, patients received these questionnaires fourteen weeks after the start of their 15-week VR program. In clinical practice, this is one week before the real discharge date and in some patients, this might even be worse if they were on holiday during the intervention period or had an extension of their training period. We suppose that this flaw yields an underestimation on the responsiveness measures in this study, because when people are in rehabilitation they cannot be at work.

Clinical recommendations

We recommend using the working status and number of working hours per week items of the iPCQ-VR to provide an estimation of short-term sick leave, which is in line with the majority of the return to work intervention studies, which use an estimate of lost time from work as their primary RTW outcome [42, 43]. A minimal important change of ≥1 working hours per week can be used for evaluative purposes for patients who are on full sick leave at baseline.

Furthermore, a minimal important change of 5.5 sick leave days per month can be considered for patients who are on full sick leave at baseline. However, this warrants caution due to the moderate AUC of 0.66. The items of the iPCQ-VR should not be used for the assessment of presenteeism. The sum of all healthcare utilization items of the TiCP-VR can be used at group level, but the single items needs further investigation. The generic item on medication use can be used at group level, but the pain-specific medication use item warrants improvement.

Conclusion

The iPCQ-VR showed good measurement properties on working status, number of hours working per week and long-term sick leave, and low measurement properties on short-term sick leave and presenteeism. The TiCP-VR showed adequate reliability on total healthcare utilization and medication use, but showed low measurement properties on the single healthcare utilization items.

Compliance with Ethical Standards

Funding

No commercial sponsorship was involved in the design and conduction of the study.

Conflict of interest

Author TB, author JvV, author CvB, author MR, and author MFD declare that they have no conflict of interest.

Ethical approval

All procedures performed were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The Medical Ethical Committee of the Academic Medical Center, Amsterdam, the Netherlands, authorized this study and decided that a full application was not required.

Informed consent

Informed consent was obtained from all individual participants included in the study.

References

1. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain. 2006 May;10(4):287-333.
2. de Vroome EM, Uegaki K, van der Ploeg CP, Treutlein DB, Steenbeek R, de Weerd M, et al. Burden of Sickness Absence Due to Chronic Disease in the Dutch Workforce from 2007 to 2011. J Occup Rehabil. 2015 Mar 25.
3. van Dongen JM, van Wier MF, Tompa E, Bongers PM, van der Beek AJ, van Tulder MW, et al. Trial-based economic evaluations in occupational health: principles, methods, and recommendations. J Occup Environ Med. 2014 Jun;56(6):563-72.
4. Hakkaart-van Roijen L, Tan S, Bouwmans C. Handleiding voor kostenonderzoek. Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg (Cost research guide. Methods and standard cost prices for economic evaluations in health care). Rotterdam: College voor zorgverzekeringen 2010.
5. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, et al. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. Eur Spine J. 2006 Mar;15 Suppl 2:S192-300.
6. Escorpizo R, Brage S, Homa D, Stucki G. Handbook of Vocational Rehabilitation and Disability Evaluation. Springer, 2014.
7. Waddell G, Burton A, Kendall N. Vocational rehabilitation: what works, for whom, and when? The Stationery Office 2013.
8. Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in healthcare settings. BMJ. 2010 Jan 18;340:c186.
9. Krol M, Brouwer W, Rutten F. Productivity costs in economic evaluations: past, present, future. Pharmacoeconomics. 2013 Jul;31(7):537-49.
10. Leggett S, van der Zee-Neuen A, Boonen A, Beaton DE, Bojinca M, Bosworth A, et al. Test-retest Reliability and Correlations of 5 Global Measures Addressing At-work Productivity Loss in Patients with Rheumatic Diseases. J Rheumatol. 2016 Feb;43(2):433-9.
11. Tang K. Estimating productivity costs in health economic evaluations: a review of instruments and psychometric evidence. Pharmacoeconomics. 2015 Jan;33(1):31-48.
12. Zhang W, Bansback N, Anis AH. Measuring and valuing productivity loss due to poor health: A critical review. Soc Sci Med. 2011 Jan;72(2):185-92.
13. Ostelo RW, de Vet HC. Clinically important outcomes in low back pain. Best Pract Res Clin Rheumatol. 2005 Aug;19(4):593-607.
14. Beaton DE, Tang K, Gignac MA, Lacaille D, Badley EM, Anis AH, et al. Reliability, validity, and responsiveness of five at-work productivity measures in patients with rheumatoid arthritis or osteoarthritis. Arthritis Care Res (Hoboken). 2010 Jan 15;62(1):28-37.
15. Bouwmans C, Krol M, Severens H, Koopmanschap M, Brouwer W, Hakkaart-van Roijen L. The iMTA Productivity Cost Questionnaire: A Standardized Instrument for Measuring and Valuing Health-Related Productivity Losses. Value Health. 2015 Sep;18(6):753-8.

16. Krol M, Brouwer W. How to estimate productivity costs in economic evaluations. Pharmacoeconomics. 2014 Apr;32(4):335-44.
17. Bouwmans C, Hakkaart-van Roijen L, Koopmanschap M, Krol M, Severens H, Brouwer W. Manual of the iMTA Productivity Cost Questionnaire (iPCQ). Rotterdam: 2013.
18. Bouwmans C, De Jong K, Timman R, Zijlstra-Vlasveld M, Van der Feltz-Cornelis C, Tan Swan S, et al. Feasibility, reliability and validity of a questionnaire on healthcare consumption and productivity loss in patients with a psychiatric disorder (TiC-P). BMC Health Serv Res. 2013;13:217.
19. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010 May;19(4):539-49.
20. Reneman MF, Beemster TT, Edelaar MJ, van Velzen JM, van Bennekom C, Escorpizo R. Towards an ICF- and IMMPACT-based pain vocational rehabilitation core set in the Netherlands. J Occup Rehabil. 2013 Dec;23(4):576-84.
21. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005 Feb;19(1):231-40.
22. Nunnally JC. Psychometric theory. 2nd ed. New York McGraw-Hill; 1978.
23. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther. 2005 Mar;85(3):257-68.
24. De Vet HCW, Terwee CB, Knol DL, Mokkink LB. Measurement in Medicine: Cambridge University Press 2011.
25. Warrens M. Weighted Kappas for 3 × 3 Tables. Journal of Probability and Statistics. Journal of Probability and Statistics. 2013 (Article ID 325831).
26. Vanbelle S. A New Interpretation of the Weighted Kappa Coefficients. Psychometrika. 2016 Jun;81(2):399-410.
27. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007 Jan;60(1):34-42.
28. Kievit AJ, Kuijer PP, Kievit RA, Sierevelt IN, Blankevoort L, Frings-Dresen MH. A reliable, valid and responsive questionnaire to score the impact of knee complaints on work following total knee arthroplasty: the WORQ. J Arthroplasty. 2014 Jun;29(6):1169-75 e2.
29. de Vet HC, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL. Clinicians are right not to like Cohen's kappa. BMJ. 2013;346:f2125.
30. Soer R, Reneman MF, Vroomen PC, Stegeman P, Coppes MH. Responsiveness and minimal clinically important change of the Pain Disability Index in patients with chronic back pain. Spine (Phila Pa 1976). 2012 Apr 15;37(8):711-5.
31. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. J Chronic Dis. 1986;39(11):897-906.
32. Kappa as a Measure of Concordance in Categorical Sorting. Accessed 20 Jun 2017; Available from: http://vassarstats.net/kappa.html.
33. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 2012 May;21(4):651-7.

34. Kessler RC, Ames M, Hymel PA, Loeppke R, McKenas DK, Richling DE, et al. Using the World Health Organization Health and Work Performance Questionnaire (HPQ) to evaluate the indirect workplace costs of illness. J Occup Environ Med. 2004 Jun;46(6 Suppl):S23-37.

35. Goetzel RZ, Hawkins K, Ozminkowski RJ, Wang S. The health and productivity cost burden of the "top 10" physical and mental health conditions affecting six large U.S. employers in 1999. J Occup Environ Med. 2003 Jan;45(1):5-14.

36. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics. 1993 Nov;4(5):353-65.

37. Ariza-Ariza R, Hernandez-Cruz B, Navarro-Compan V, Leyva Pardo C, Juanola X, Navarro-Sarabia F. A comparison of telephone and paper self-completed questionnaires of main patient-related outcome measures in patients with ankylosing spondylitis and psoriatic arthritis. Rheumatol Int. 2013 Nov;33(11):2731-6.

38. Bushnell DM, Reilly MC, Galani C, Martin ML, Ricci JF, Patrick DL, et al. Validation of electronic data capture of the Irritable Bowel Syndrome--Quality of Life Measure, the Work Productivity and Activity Impairment Questionnaire for Irritable Bowel Syndrome and the EuroQol. Value Health. 2006 Mar-Apr;9(2):98-105.

39. Zhang W, Bansback N, Kopec J, Anis AH. Measuring time input loss among patients with rheumatoid arthritis: validity and reliability of the Valuation of Lost Productivity questionnaire. J Occup Environ Med. 2011 May;53(5):530-6.

40. Anema JR, Schellart AJ, Cassidy JD, Loisel P, Veerman TJ, van der Beek AJ. 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 Dec;19(4):419-26.

41. Pole JD, Franche RL, Hogg-Johnson S, Vidmar M, Krause N. Duration of work disability: a comparison of self-report and administrative data. Am J Ind Med. 2006 May;49(5):394-401.

42. Cullen KL, Irvin E, Collie A, Clay F, Gensby U, Jennings PA, et al. Effectiveness of Workplace Interventions in Return-to-Work for Musculoskeletal, Pain-Related and Mental Health Conditions: An Update of the Evidence and Messages for Practitioners. J Occup Rehabil. 2018 Mar;28(1):1-15.

43. Vogel AP, Barker SJ, Young AE, Ruseckaite R, Collie A. What is return to work? An investigation into the quantification of return to work. Int Arch Occup Environ Health. 2011 Aug;84(6):675-82.

Appendix 1. iPCQ-VR

No.	English version	Dutch version	Answer options	iPCQ-VR	Original iPCQ
1	Are you working full-time at this moment?	Bent u op dit moment volledig aan het werk?	1=Yes, 2=No I am partly at work, 3=No I am on 100% sick leave	x	
2	If *"No, I am partly at work" for No. 1:* How many hours are you working per week at this moment?	Hoeveel uur per week werkt u op dit moment?	Continuous (0-40)	x	
3	Were you on sick leave in the past 4 weeks?	Bent u in de afgelopen 4 weken afwezig geweest van uw werk omdat u ziek was?	1=Yes, 2=No	x	x
4	If *"Yes" for No. 3:* I was ··· days sick from work. (Only count working days in the past 4 weeks)	Ik ben ··· dagen afwezig geweest. (Tel alleen de werkdagen in de afgelopen 4 weken)	Continuous (0-31)	x	x
5	If *"Yes" for No. 3:* What was the reason for your sick leave?	Wat was de reden van uw ziekmelding?	1=Sick leave related to subacute or chronic musculoskeletal pain, 2=Other reasons, such as flu	x	
6	Did you have a consecutive period of sick leave longer than 4 weeks?	Was u langer dan de gehele periode van 4 weken afwezig van uw werk doordat u ziek was? Het gaat om een aaneengesloten periode van werkverzuim.	1=Yes, 2=No	x	x
7	If *"Yes" for No. 6:* When did you report yourself sick?	Wanneer heeft u zich ziek gemeld?	dd-mm-yyyy	x	x
8	If *"Yes" for No. 6:* What was the reason for the sick leave?	Wat was de reden van uw ziekmelding?	1=Sick leave related to subacute or chronic musculoskeletal pain, 2=Other reasons, such as flu	x	
9	Were you less productive at your work in the past 4 weeks due to mental or physical problems?	Waren er in de afgelopen 4 weken dagen waarop u wel gewerkt heeft, maar tijdens uw werk last had van lichamelijke of psychische problemen?	1=Yes, 2=No	x	x

Appendix 1. iPCQ-VR (Continued)

10	If *"Yes" for No. 9:* On how many working days were you less productive in the past 4 weeks due to physical or mental problems?	Op hoeveel werkdagen had u tijdens uw werk last van uw lichamelijke of psychische problemen? Tel alleen de werkdagen in de afgelopen 4 weken.	Continuous (0-31)	x	x
11	If *"Yes" for No. 9:* Give a presenteeism score	Op de dagen dat u last had, kon u misschien niet zoveel werk doen als normaal. Hoeveel werk kon u op deze dagen gemiddeld doen? Kijk naar de cijfers hieronder. Een 10 betekent dat u op deze dagen net zoveel kon doen als normaal. Een 0 betekent dat u op deze dagen niets kon doen.	0=I couldn't do anything, 10=I could do the same as normal	x	x

Appendix 2. TiCP-VR

No.	English version	Dutch version	Answer options	Mandatory	Optional
01A	Did you consult with a General Practitioner at any time during the past four weeks?	Hebt u in de afgelopen 4 weken contact gehad met de HUISARTS? Zo ja, hoeveel keer?	1=Yes, 2=No	x	
01B	Number of visits on appointment	Aantal bezoeken op afspraak	··· times		x
01C	Number of consulting hours	Aantal spreekuurbezoeken	··· times		x
01D	Number of telephone consultations	Aantal telefonische consulten	··· times		x
01E	Number of home visits	Aantal huisbezoeken	··· times		x
	Were these consultations due to your pain problems for which you have been referred to vocational rehabilitation?	Waren deze contacten vanwege de klachten waarvoor u voor (arbeids)revalidatie bent verwezen?	1=Yes, 2=No		x
01G	How many consultations were due to your pain problems for which you have been referred to vocational rehabilitation?	Hoe vaak hadden deze contacten betrekking op de klachten waarvoor u bent verwezen voor (arbeids)revalidatie?	··· times		x
02A	Did you consult with a Physical therapist, Occupational therapist, Exercise therapist or Manual therapist at any time during the past four weeks?	Hebt u in de afgelopen 4 weken contact gehad met een FYSIOTHERAPEUT, OEFENTHERAPEUT, ERGOTHERAPEUT OF MANUEEL THERAPEUT? Zo ja, hoeveel keer? (met oefentherapeut wordt oefentherapeut Mensendieck of oefentherapeut Ceasar bedoeld)	1=Yes, 2=No	x	
02B	Physical therapist	Fysiotherapeut	··· times		x
02C	Exercise therapist	Oefentherapeut	··· times		x
ZG02D	Occupational therapist	Ergotherapeut	··· times		x
ZG02E	Manual therapist	Manueel therapeut	··· times		x
02F	Were these consultations due to your pain problems for which you have been referred to vocational rehabilitation?	Waren deze contacten vanwege de klachten waarvoor u voor (arbeids)revalidatie bent verwezen?	1=Yes, 2=No		x

Appendix 2. (Continued)

No.	English version	Dutch version	Answer options	Mandatory	Optional
02G	How many consultations were due to your pain problems for which you have been referred to vocational rehabilitation?	Hoe vaak hadden deze contacten betrekking op de klachten waarvoor u bent verwezen voor (arbeids)revalidatie?	··· times		x
03A	Did you consult with an Occupational Physician at any time during the past four weeks?	Hebt u in de afgelopen 4 weken contact gehad met een BEDRIJFSARTS? Zo ja, hoeveel keer?	1=Yes, 2=No	x	
03B	Yes, namely	Ja, namelijk	··· times		x
03C	Were these consultations due to your pain problems for which you have been referred to vocational rehabilitation?	Waren deze contacten vanwege de klachten waarvoor u voor (arbeids)revalidatie bent verwezen?	1=Yes, 2=No		x
03D	How many consultations were due to your pain problems for which you have been referred to vocational rehabilitation?	Hoe vaak hadden deze contacten betrekking op de klachten waarvoor u bent verwezen voor (arbeids)revalidatie?	··· times		x
04A	Did you consult with a Reintegration advisor, Reintegration specialist, Occupational specialist or Job coach at any time during the past four weeks?	Hebt u in de afgelopen 4 weken contact gehad met een RE-INTEGRATIEADVISEUR, RE-INTEGRATIEDESKUNDIGE, ARBEIDSDESKUNDIGE of JOBCOACH? Zo ja, hoeveel keer?	1=Yes, 2=No	x	
04B	Reintegration advisor	Re-integratieadviseur	··· times		x
04C	Reintegration specialist	Re-integratiedeskundige	··· times		x
04D	Occupational specialist	Arbeidsdeskundige	··· times		x
04E	Jobcoach	Jobcoach	··· times		x
04F	Were these consultations due to your pain problems for which you have been referred to vocational rehabilitation?	Waren deze contacten vanwege de klachten waarvoor u voor (arbeids)revalidatie bent verwezen?	1=Yes, 2=No		x

Appendix 2. (Continued)

No.	English version	Dutch version	Answer options	Mandatory	Optional
04G	How many consultations were because of your pain problems for which you are referred to vocational rehabilitation?	Hoe vaak hadden deze contacten betrekking op de klachten waarvoor u bent verwezen voor (arbeids)revalidatie?	··· times		x
05A	Did you consult with a Insurance Physician at any time during the past four weeks?	Hebt u in de afgelopen 4 weken contact gehad met een VERZEKERINGSARTS? Zo ja, hoeveel keer?	1=Yes, 2=No	x	
05B	Yes, namely	Ja, namelijk			x
05C	Were these consultations due to your pain problems for which you have been referred to vocational rehabilitation?	Waren deze contacten vanwege de klachten waarvoor u voor (arbeids)revalidatie bent verwezen?	1=Yes, 2=No		x
05D	How many consultations were due to your pain problems for which you have been referred to vocational rehabilitation?	Hoe vaak hadden deze contacten betrekking op de klachten waarvoor u bent verwezen voor (arbeids)revalidatie?			x
06A	Were there any adaptations to your workplace at any time during the past four weeks?	Hebt u in de afgelopen 4 weken AANPASSINGEN AAN UW WERKPLEK gehad?	1=Yes, 2=No	x	
06B	Which workplace adaptations were applied?	Welke aanpassing(en) heeft u aan uw werkplek gehad?	string		x
06C	Were these workplace adaptations due to your pain problems for which you have been referred to vocational rehabilitation?	Was deze aanpassing / waren deze aanpassingen vanwege de klachten waarvoor u voor (arbeids)revalidatie bent verwezen?	1=Yes, 2=No		x
07A	Did you consult with a Medical Specialist at a hospital at any time during the past four weeks?	Hebt u in de afgelopen 4 weken contact gehad met een MEDISCH SPECIALIST OP DE POLIKLINIEK VAN EEN ZIEKENHUIS zonder dat u was opgenomen in een ziekenhuis? Zo ja, hoeveel keer? (Voorbeelden van medische specialisten zijn een cardioloog, een reumatoloog, ee	1=Yes, 2=No	x	

Appendix 2. (Continued)

No.	English version	Dutch version	Answer options	Mandatory	Optional
07B	Yes, namely		1=Orthopedic surgeon, 2=Surgeon, 3=Gynecologist, 4=Internist, 5=Urologist, 6=Rheumatologist, 7=Anesthesiologist, 8=Neurologist, 9=Other		x
07C	Specify other	Specificeer anders			x
07D	Per Medical Specialist: Number of consultations at any time during the past four weeks.		··· times		x
07E	Were these consultations due to your pain problems for which you have been referred to vocational rehabilitation?	Waren deze contacten vanwege de klachten waarvoor u voor (arbeids)revalidatie bent verwezen?	1=Yes, 2=No		x
07F	How many consultations were due to your pain problems for which you have been referred to vocational rehabilitation?	Hoe vaak hadden deze contacten betrekking op de klachten waarvoor u bent verwezen voor (arbeids)revalidatie?	··· times		x
08A	Have you stayed in a healthcare setting at any time during the past four weeks?	Bent u in de afgelopen 4 weken OPGENOMEN geweest in een instelling binnen de gezondheidszorg? Zo ja, hoeveel dagen (d.w.z. één nacht of meer in bijvoorbeeld ziekenhuis, revalidatiecentrum, psychiatrisch ziekenhuis.)	1=Yes, 2=No	x	
08B	Yes, namely	Ja, namelijk	1=Psychiatric hospital, 2=Rehabilitation center, 3=Hospital (general), 4=Hospital (Acadamic), 5=Another health care setting		x
08C	Specify other	Specificeer anders			x

Appendix 2. (Continued)

No.	English version	Dutch version	Answer options	Mandatory	Optional
08D	Number of days at any time during the past four weeks	Aantal dagen in de afgelopen 4 weken	··· times		x
08E	Were these stays due to your pain problems for which you have been referred to vocational rehabilitation?	Was deze opname/ waren deze opnames vanwege de klachten waarvoor u voor (arbeids)revalidatie bent verwezen?	1=Yes, 2=No		x
08F	How many consultations were due to your pain problems for which you have been referred to vocational rehabilitation?	Hoe vaak hadden deze opnames betrekking op de klachten waarvoor u bent verwezen voor (arbeids)revalidatie?	··· times		x
09A	Did you consult with a Psychiatrist or Psychologist at any time during the past four weeks?	Hebt u in de afgelopen 4 weken contact gehad met een PSYCHIATER, PSYCHOLOOG?	1=Yes, 2=No	x	
09B	Yes, namely	Ja, namelijk	··· times		x
09C	If yes, in which setting was your consultation?	Zo ja, in welke SETTING vond dit contact plaats?	1=personal (group)practice, 2=Mental Health Care*, 3=Rehabilitation center, 4=Hospital (general), 5=Hospital (academic), 6=Psychiatric hospital, 7=Other		x
09D	Specify other	Specificeer anders	string		x
09E	Were these consultations due to your pain problems for which you have been referred to vocational rehabilitation?	Waren deze contacten vanwege de klachten waarvoor u voor (arbeids)revalidatie bent verwezen?	1=Yes, 2=No		x
09F	How many consultations were due to your pain problems for which you have been referred to vocational rehabilitation?	Hoe vaak hadden deze contacten betrekking op de klachten waarvoor u bent verwezen voor (arbeids)revalidatie?	··· times		x

Appendix 2. (Continued)

No.	English version	Dutch version	Answer options	Mandatory	Optional
10A	Did you consult with a Social worker at any time during the past four weeks?	Hebt u in de afgelopen 4 weken contact gehad met een MAATSCHAPPELIJK WERK(ST)ER? Zo ja, hoeveel keer?	1=Yes, 2=No	x	
10B	Yes, namely	Ja, namelijk	··· times		x
10C	Were these consultations due to your pain problems for which you have been referred to vocational rehabilitation?	Waren deze contacten vanwege de klachten waarvoor u voor (arbeids)revalidatie bent verwezen?	1=Yes, 2=No		x
10D	How many consultations were due to your pain problems for which you have been referred to vocational rehabilitation?	Hoe vaak hadden deze contacten betrekking op de klachten waarvoor u bent verwezen voor (arbeids)revalidatie?	··· times		x
11A	Did you consult with a Dietician at any time during the past four weeks?	Hebt u in de afgelopen 4 weken contact gehad met een DIETIST? Zo ja, hoeveel keer?	1=Yes, 2=No	x	
11B	Yes, namely	Ja, namelijk	··· times		x
11C	Were these consultations due to your pain problems for which you have been referred to vocational rehabilitation?	Waren deze contacten vanwege de klachten waarvoor u voor (arbeids)revalidatie bent verwezen?	1=Yes, 2=No		x
11D	How many consultations were due to your pain problems for which you have been referred to vocational rehabilitation?	Hoe vaak hadden deze contacten betrekking op de klachten waarvoor u bent verwezen voor (arbeids)revalidatie?	··· times		x
12A	Did you receive home care at any time during the past four weeks?	Hebt u in de afgelopen 4 weken gebruik gemaakt van THUISZORG? Zo ja, hoeveel uur? (Wilt u alle uren in de afgelopen 4 weken bij elkaar optellen?)	1=Yes, 2=No	x	
12B	Yes, namely	Ja, namelijk	··· times		x
12C	Was this use of home care due to your pain problems for which you have been referred to vocational rehabilitation?	Was dit gebruik van thuiszorg vanwege de klachten waarvoor u voor (arbeids)revalidatie bent verwezen?	1=Yes, 2=No		x

Appendix 2. (Continued)

No.	English version	Dutch version	Answer options	Mandatory	Optional
12D	How many hours were due to your pain problems for which you have been referred to vocational rehabilitation?	Hoeveel uur had deze thuiszorg betrekking op de klachten waarvoor u bent verwezen voor (arbeids)revalidatie?	··· hours		x
13A	Did you consult with a practitioner of Alternative Medicine at any time during the past four weeks?	Hebt u in de afgelopen 4 weken contact gehad met een ALTERNATIEVE GENEZER? Zo ja, hoeveel keer? (Voorbeelden van een alternatieve genezer zijn o.a. homeopaat, acupuncturist, natuurgenezer, haptonoom, chiropractor, iriscopist, osteopaat, etc)	1=Yes, 2=No	x	
13B1	Name alternative care	Naam alternatieve therapie:	string		x
13B2	Number of consultations at any time during the past four weeks.	Alternatieve therapie: Aantal contacten in de afgelopen 4 weken	··· times		x
13C1	Name alternative care 2:	Naam alternatieve therapie 2:	string		x
13C2	Alternative care 2: Number of consultations at any time during the past four weeks.	Alternatieve therapie 2: Aantal contacten in de afgelopen 4 weken	··· times		x
13D1	Name alternative care 3:	Naam alternatieve therapie 3:	string		x
13D2	Alternative care 3: Number of consultations at any time during the past four weeks.	Alternatieve therapie 3: Aantal contacten in de afgelopen 4 weken	··· times		x
13E	Were these consultations due to your pain problems for which you have been referred to vocational rehabilitation?	Waren deze contacten vanwege de klachten waarvoor u voor (arbeids)revalidatie bent verwezen?	1=Yes, 2=No		x
13F	How many consultations were due to your pain problems for which you have been rehabilitation?	Hoe vaak hadden deze contacten betrekking op de klachten waarvoor u bent verwezen voor (arbeids)revalidatie?	··· times		x

Appendix 2. (Continued)

No.	English version	Dutch version	Answer options	Mandatory	Optional
14A	Did you used Medication at any time during the past four weeks?	Hebt u in de afgelopen 4 weken MEDICIJNEN gebruikt? (Pakt u zo mogelijk de verpakking van de medicatie erbij. Medicijnen tijdens ziekenhuisopname NIET meerekenen, evenmin als de anticonceptie pil.)	1=Yes, 2=No	x	
14B	Was this medication use due to your pain problems for which you have been referred to vocational rehabilitation?	Was dit medicijngebruik vanwege de klachten waarvoor u voor (arbeids)revalidatie bent verwezen?	1=Yes, fully, 2=Yes, partly, 3=No		x
14B1	Name drug 1	Naam medicijn 1:	string		x
14B2	Dose	Dosis			x
14B3	Number of times per day	Aantal keer per dag:	··· times		x
14B4	Number of days in the past four weeks	Aantal dagen in de afgelopen 4 weken:	··· times		x
	The same as ZG14B1-ZG14B4 up to eight medications.				x
15	Maybe you have some questions or remarks? If so, you can write them here.	Misschien heeft u nog vragen of opmerkingen? Schrijft u deze dan hiernaast op.	string		x

Appendix 3. 2 x 2 contingency tables iPCQ-VR

Total sample (stable and unstable participants)					Stable group of participants			
Are you working full-time at this moment?								
	yes	part-time at work	100% sick leave			yes	part-time at work	100% sick leave
yes	7	0	0		yes	7	0	0
part-time at work	1	23	5		part-time at work	0	20	1
100% sick leave	0	1	13		100% sick leave	0	0	13
Were you on sick leave in the past 4 weeks?								
		no	yes			no	yes	
	no	19	6		no	18	5	
	yes	8	17		yes	8	16	
What was the reason for your sick leave?								
		2	1			2	1	
	2	2	0		2	2	0	
	1	1	14		1	1	13	
1=Sick leave related to subacute or chronic musculoskeletal pain, 2=Other reasons such as flu								
Did you have a consecutive period of sick leave longer than 4 weeks?								
		no	yes					
	no	29	3			not applicable		
	yes	3	15					
What was the reason for the sick leave?								
		2	1					
	2	1	0			not applicable		
	1	0	14					
1=Sick leave related to subacute or chronic musculoskeletal pain, 2=Other reasons such as flu								
Were you less productive at your work in the past 4 weeks, due to mental or physical problems?								
		no	yes			no	yes	
	no	9	4		no	9	4	
	yes	8	29		yes	7	22	

Appendix 4. 2 x 2 contingency table TiCP-VR

Total sample (stable and unstable participants)					Stable group of participants			
				No.				
Did you consult with a General Practitioner at any time during the past four weeks?								
		no	yes			no	yes	
	no	18	6	01A	no	17	6	
	yes	14	12		yes	14	9	
Did you consult with a Physical therapist, Occupational therapist, Exercise therapist or Manual therapist at any time during the past four weeks?								
		no	yes			no	yes	
	no	16	0	02A	no	16	0	
	yes	11	23		yes	10	16	
Did you consult with an Occupational Physician at any time during the past four weeks?								
		no	yes			no	yes	
	no	7	6	03A	no	6	4	
	yes	18	19		yes	15	18	
Did you consult with a Reintegration advisor, Reintegration specialist, Occupational specialist or Job coach at any time during the past four weeks?								
		no	yes			no	yes	
	no	33	6	04A	no	32	5	
	yes	1	9		yes	1	6	
Did you consult with an Insurance Physician at any time during the past four weeks?								
		no	yes			no	yes	
	no	47	1	05A	no	47	0	
	yes	2	0		yes	2	0	
Were there any adaptations to your workplace at any time during the past four weeks?								
		no	yes			no	yes	
	no	35	4	06A	no	35	3	
	yes	4	7		yes	4	6	
Did you consult with a Medical Specialist at a hospital at any time during the past four weeks?								
		no	yes			no	yes	
	no	33	3	07A	no	31	3	
	yes	6	8		yes	6	6	
Did you stay in a healthcare setting at any time during the past four weeks?								
		no	yes			no	yes	
	no	50	0	08A	no	50	0	
	yes	0	0		yes	0	0	

Appendix 4. (Continued)

Total sample (stable and unstable participants)					Stable group of participants			
				No.				
Did you consult with a Psychiatrist or Psychologist at any time during the past four weeks?								
		no	yes			no	yes	
	no	31	10	09A	no	31	9	
	yes	2	7		yes	2	5	
Did you consult with a Social worker at any time during the past four weeks?								
		no	yes			no	yes	
	no	48	1	10A	no	48	1	
	yes	0	0		yes	0	0	
Did you consult with a Dietician at any time during the past four weeks?								
		no	yes			no	yes	
	no	48	0	11A	no	48	0	
	yes	0	1		yes	0	1	
Did you receive home care at any time during the past four weeks?								
		no	yes			no	yes	
	no	48	1	12A	no	48	1	
	yes	0	0		yes	0	0	
Did you consult witha practitioner of Alternative Medicine at any time during the past four weeks?								
		no	yes			no	yes	
	no	41	1	13A	no	41	1	
	yes	3	4		yes	3	3	
Did you use Medication at any time during the past four weeks?								
		no	yes			no	yes	
	no	13	2	14A	no	13	1	
	yes	3	30		yes	3	21	
Was these medication use due to your pain problems for which you have been referred to vocational rehabilitation?								
	no	yes, fully	yes, partly			no	yes, fully	yes, partly
no	2	0	0		no	2	0	0
yes, fully	1	12	4	14B	yes, fully	1	10	4
yes, partly	1	5	5		yes, partly	1	5	4
Health care usage pain-specific (sum)								
		no	yes			no	yes	
	no	4	4	total	no	3	2	
	yes	13	74		yes	7	70	

CHAPTER 4

The interpretation of change score of the pain disability index after vocational rehabilitation is baseline dependent

T. Beemster
C. van Bennekom
J. van Velzen
M. Reneman
M. Frings-Dresen

Health Qual Life Outcomes. 2018 Sep 14;16(1):182

Abstract

Background

The Pain Disability Index (PDI) is a widely-used instrument to measure pain-related disability. The aim of this study was to assess the responsiveness and interpretation of change score of the PDI in patients with chronic musculoskeletal pain (CMP) at discharge of vocational rehabilitation.

Methods

Retrospective data of patients with CMP who attended vocational rehabilitation between 2014-2017 was used. The anchor-based method was used to assess the responsiveness of the total sample and of PDI baseline quartile groups. A receiver operating characteristic curve was performed, including Area Under the Curve (AUC) and Minimal Important Change (MIC).

Results

The PDI showed responsive to detect clinically relevant changes in pain-related disability at discharge of vocational rehabilitation (AUC 0.79). A PDI change score of 13 points (MIC 12.5) can be considered as a real change in pain-related disability for the total study sample, and a PDI change score of 7-20 points can be considered as a real change in pain-related disability for PDI lowest and highest baseline quartile scores.

Conclusion

The PDI is responsive in patients with CMP at discharge of vocational rehabilitation. The interpretation of change score depends on PDI baseline score. Patients with a PDI baseline score of ≤27 should decrease minimal 7 points, patients with a baseline score between 28-42 should decrease minimal 15 points, and patients with a baseline score ≥43 should decrease minimal 20 points.

Keywords

Clinical relevance, Minimal Important Difference, Pain Disability Index, occupational rehabilitation, interpretation of change, chronic pain.

Background

Chronic Musculoskeletal Pain (CMP) negatively affects quality of life, daily activities and social and working lives [1]. A decrease of pain-related disability is a desired outcome measure after rehabilitation for people with CMP [2]. A widely used and studied instrument to measure pain-related disability is the Pain Disability Index (PDI) [2, 3]. The PDI is a generic instrument: it can be administered to different patient groups, for example, chronic low back pain, fibromyalgia, cancer, or chronic widespread pain. The PDI is a valid [4-6] and reliable [6, 7] instrument. The utility of the PDI is high because it is easy to comprehend, it can be administered in a very short time, and it consists of only 7 questions [8].

However, the responsiveness, measurement error, and interpretability of change score of the PDI have scarcely been addressed. Responsiveness is the ability of a questionnaire to detect clinically important changes over time (for example, at discharge of a rehabilitation program) [9]. An outcome instrument should be able to distinguish clinically important change from measurement error [10]. The relation between responsiveness and measurement error should be made to interpret the (change) score of a questionnaire [10]. Nevertheless, to our knowledge, only one study [8] has assessed responsiveness and one other study [6] has assessed measurement error of the PDI. Good responsiveness (Area Under the Curve (AUC) of 0.76) was found in patients with chronic low back pain at discharge of a pain rehabilitation program in the Netherlands, and a minimal important change (MIC) of 8.5-9.5 points (depending on which anchor was used) was calculated [8]. In addition, a MIC value of 9.5 means that a decrease in PDI score of 9.5 points or more is a clinically meaningful improvement in pain-related disability. Measurement error, expressed in the Smallest Detectable Change (SDC), of 17.9 points was found in a sample with acute back pain, chronic low back pain, and widespread pain [6]. However, a connection between the MIC and the SDC (which refers to the interpretation of change score of the PDI), respectively, was not provided in the aforementioned studies. If we combine the MIC of 9.5 with the SDC of 17.9, we conclude that the PDI is responsive to change in patients with chronic back pain, but that it is uncertain if these are 'real' changes or are due to measurement error [11].

The aforementioned studies on responsiveness and measurement error were performed with patients attending pain rehabilitation in the Netherlands. It is unknown, however, what the responsiveness and interpretation of change score

of the PDI is for patients at discharge of vocational rehabilitation (VR). Vocational rehabilitation is a “multi-professional evidence-based approach” that is provided in different settings, services, and activities to working age individuals with health-related impairments, limitations, or restrictions with work functioning, and whose primary aim is to “optimize work participation” [12]. However, it can be expected that the majority of patients referred to VR have paid work. In contrast, in pain rehabilitation samples, less than 50% of the patients have paid work [6, 13]. Since work is generally good for physical and mental health and well-being, and unemployment is associated with poorer physical and mental health and well-being [14], we expect that patients referred to VR are less disabled (i.e. lower PDI score) compared to patients referred to pain rehabilitation. We therefore assume that there is less room for improvement compared to patients with more severe pain-related disability and that this could result in lower MIC and change scores. This has, however, not yet been studied. Therefore, the aim of this study is to assess the responsiveness and interpretation of change score of the PDI in patients with chronic musculoskeletal pain at discharge of vocational rehabilitation.

Methods

The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist was applied in the design of the study [9, 15, 16].

Study sample

The study sample consisted of CMP patients who attended vocational rehabilitation (VR) between November 2014 and July 2017 in the Netherlands. Vocational rehabilitation is a multidisciplinary bio-psychosocial group-based program for workers with CMP and decreased work participation. The VR program is described in detail elsewhere [17]. The study sample was derived from seven vocational rehabilitation centers in the Netherlands. These seven centers are part of a nationwide network in the Netherlands and the outline and content of VR is similar at each center. The inclusion criteria for attending VR were: 1) being of working age (18 to 65 years); 2) suffering from subacute (6 to 12 weeks) or chronic (>12 weeks) nonspecific musculoskeletal pain; 3) decreased work participation (i.e. part-time or full-time sick leave or reduced productivity while at work). The exclusion criteria were: 1) not motivated to participate in the multidisciplinary group-based program; 2) psychiatric disorders; 3) physical disorders with the expectation that tissue and function recovery will take place at normal rates; and 4) conflict situations with employer. Extra inclusion criteria for this study were: 1) being able to complete questionnaires in Dutch; and 2) having completed the Pain Disability Index at baseline and discharge of VR.

Procedures

Data were collected using a core set of standardized web-based patient-reported questionnaires [18]. For this study, we only used the questionnaires on sample characteristics, including Pain Disability Index, assessed at baseline (T0) and discharge (T1); and Global Perceived Effect, assessed at T1 only. At T0 and T1, patients received an email with login data and the request to complete questionnaires (at home) on a website. Baseline questionnaires were sent to patients 1-2 weeks before a multidisciplinary screening, and the discharge questionnaires were sent to patients one week before discharge date. Because this study contains routinely collected and anonymous data of care as usual programs, the Medical Ethical Committee of the Academic Medical Center,

Amsterdam, the Netherlands, authorized this study and decided that a full application was not required (reference number: A1 17.405).

Outcome instrument: The Pain Disability Index

The Pain Disability Index (PDI) is a 7-item questionnaire to investigate the magnitude of self-reported pain-related disability, independent from region of pain or pain-related diagnosis. The items of the questionnaire are assessed on a 0-10 numeric rating scale in which 0 means no disability and 10 is maximum disability. The sum of the seven items equals the total score of the PDI, which ranges from 0 to 70, with higher scores reflecting higher interference of pain with daily activities. The PDI measures family / home responsibilities, recreation, social activity, occupation, sexual behavior, self-care and life support activity [3]. Missing items were resolved as follows: patients were allowed to miss no more than 1 question on the PDI. In this case, the missing value was replaced by the patient cluster mean. As the PDI only consists of 7 questions, the patient was excluded from the study [6] if the patient missed more than one question on the PDI.

Anchor: global perceived effect of treatment

A global perceived effect (GPE) item was used as the anchor (external criterion) in this study. An anchor is a global rating scale in which patients are asked, in a single question at follow-up, to indicate how much their pain has changed since baseline [19]. The pain anchor was assessed as follows: 'How are your (pain) complaints at this moment compared to pre-treatment?'. The anchor was assessed on a 7-point Likert scale: extremely worsened, much worsened, little worsened, unchanged, little improved, much improved, completely improved.

Data analyses

Responsiveness

Responsiveness in this study was defined as the ability of the PDI to detect clinically relevant changes in pain-related disability at discharge of vocational rehabilitation [9]. To calculate responsiveness we used the anchor-based receiver operating characteristics (ROC) method [20]. Sensitivity and specificity for change plotted by receiver operating characteristics (ROC) curve and Area Under the Curve (AUC) were calculated [10]. The AUC is the probability of correctly discriminating between improved and unchanged patients. When the

AUC was more than 0.70, responsiveness was considered sufficient [10]. Minimal Important Change (MIC) was measured by determining the optimal cut-off point, i.e. the point where the sum of sensitivity and 1-specificity was maximal. Sensitivity and specificity range from 0 to 1.00, where higher numbers reflect higher sensitivity or specificity. Because the objective of the responsiveness analysis was to differentiate between improved and unchanged patients, the anchor scores were dichotomized into a subgroup with the score "improved" (much improved and completely improved) and a subgroup with the score "unchanged" (little worsened, unchanged and little improved) [8]. The group with the score "worsened" (much worsened and extremely worsened) was not included in the analyses (n=14). We used the improved and unchanged groups to calculate the MIC [10, 20].

Baseline-dependent analyses

In a secondary analysis we stratified the analysis on PDI baseline quartile scores, to assess whether the level of pain-related disability on baseline had a modifying effect on the MIC. Based on earlier research [21, 22] we hypothesized that higher PDI scores at baseline (that is, more disabled patients thus higher PDI score) had more room for improvement, including higher change scores and MIC values compared to patients with lower baseline scores.

Floor and ceiling effects

Floor or ceiling effects were considered to be present if more than 15% of the respondents achieved the lowest or highest possible score (0-70, respectively) [10]. We gave a positive rating for (the absence of) floor and ceiling effects if no floor or ceiling effects were present in the PDI baseline quartiles [10].

Measurement error

Measurement error was analyzed by calculating the Standard Error of Measurement (SEM=SD$\sqrt{1}$-ICC) [23]. The SD was determined from an ANOVA analysis with the formula ($\sqrt{(SStotal\ /(n-1))}$) [10, 23]. As proposed by Terwee et al. [11], we derived the SD from our study sample for the patients with a non-significant change in PDI score (PDI total score T1 – PDI total score T0 = p>0.05). Independent samples T test showed a non-significant change in PDI score when the PDI change score ranged from -6 to +6. The ICC of the SEM formula was obtained from a study with a similar study sample [6]. In a next step, the SEM was converted into the smallest detectable changes at individual level (SDCindividual=1.96*$\sqrt{2}$*SEM). This number reflects the smallest within-person change in a score that can be considered to be a real change above any

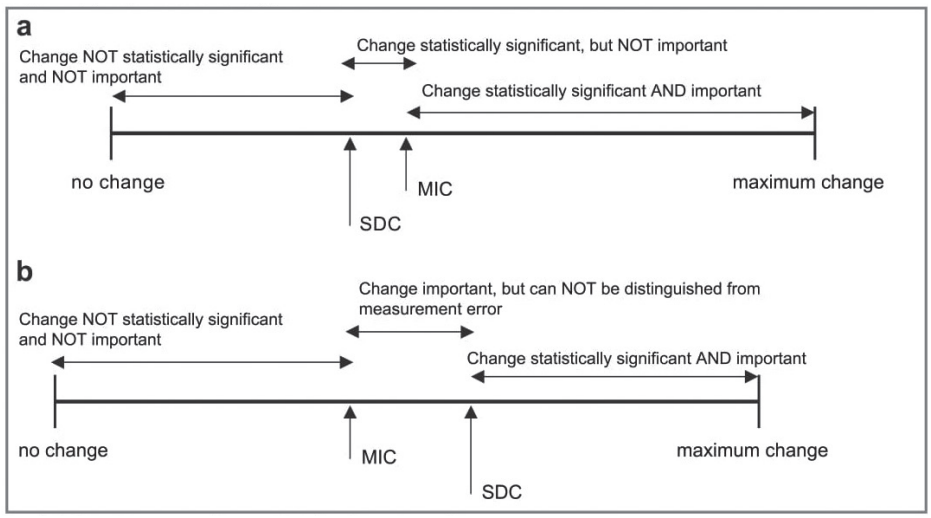
measurement error within one individual. In the final step, the SDC individual was converted into the smallest detectable change for a group (SDC group) by dividing SDC individual by $\sqrt{n}$.

Interpretability

Interpretability is defined as the degree to which one can assign qualitative meaning to quantitative scores [10]. To enhance interpretability, we will present baseline scores and change scores of various (sub)groups. For the interpretability of change scores, we calculated mean changes and 95% confidence intervals of mean changes of the total study sample and of the PDI baseline quartiles. We gave a positive rating for a real change in decrease of pain-related disability when the PDI change score was larger than the SDC, and if the SDC was smaller than the MIC [10, 19] (see Figure 1).

All analyses were performed using SPSS 23 for Windows (SPSS Inc., Chicago, USA). The demographic data of the individuals were described by means and standard deviations (SD), or inter-quartile range in the case of no normal distribution. The assumption of normal data distribution was visually verified using histograms and QQ-plots.

Figure 1. Interpretation of PDI change scores



MIC, Minimal Important Change; SDC, Smallest Detectable Change
Figure obtained from Terwee et al. [10]

Results

A total of 341 patients completed the PDI questionnaire on baseline and discharge. Mean age was 46.5 (±10.9) years, and 57% of the patients were woman. 91% of the patients were employed and 63% were on sick leave in the preceding month prior to baseline measurement. Patients suffered from 3.4 (±2.4) pain locations, which were located in the back (76%), lower extremities (35%) and upper extremities (29%). 74% had pain complaints for longer than six months. The average pain score was 5.4 (±2.3), the worst pain score was 6.8 (±2.5) and the PDI mean score was 34.7 (±11.7). Mean duration between baseline questionnaires and the start of VR was 8±4.4 weeks and mean duration between the start of VR and completion of the discharge questionnaires was 15±1.1 weeks. Table 1 shows all background characteristics of the study sample.

Responsiveness

The responsiveness parameters (AUC, MIC, sensitivity and specificity) of the total study sample and the baseline quartile scores are presented in Table 2, and the corresponding ROC curves are presented in Figure 2. The AUC of the total sample was 0.79 (0.74-0.84), with a sensitivity of 0.68, a specificity of 0.73, and a corresponding MIC of 12.5 (Fig 2a). The AUC of PDI baseline quartile 1 was 0.70 (0.59-0.81), with a sensitivity of 0.68, a specificity of 0.67, and a corresponding MIC of 6.5. The AUC of PDI baseline quartile 2 was 0.87 (0.79-0.95), with a sensitivity of 0.81, a specificity of 0.80, and a corresponding MIC of 14.5. The AUC of PDI baseline quartile 3 was 0.83 (0.73-0.93), with a sensitivity of 0.71, a specificity of 0.73, and a corresponding MIC of 14.5. The AUC of PDI baseline quartile 4 was 0.85 (0.77-0.93), with a sensitivity of 0.79, a specificity of 0.81, and a corresponding MIC of 19.5. In summary, the mean AUC of the total sample and of all PDI quartiles was sufficient, and only for quartile 1 the 95% confidence interval of the AUC felt bellow the cut off of 0.70, indicating slightly insufficient responsiveness for this quartile (also indicated by the shape of the ROC curve (Fig 2b)).

Table 1. Characteristics of the study sample

	Unit of measurement	Vocational rehabilitation (n=351)
Age (years)	Mean (sd)	46.5 (10.9)
Gender (female)	%	57.1
Education		
Low	%	15.1
Medium	%	54.0
High	%	24.9
Other	%	6.0
Work situation		
Employed	%	90.6
Student	%	0.6
Benefit	%	2.6
Other	%	6.3
Sick leave in the past month (yes)	%	63.4
Number of pain locations [a-c]	Mean (sd)	3.4 (2.4)
	Median (IQR)	3 (1-5)
Pain location		
Spine (yes) [a]	%	76.1
Lower extremities (yes) [b]	%	35.0
Upper extremities (yes) [c]	%	29.1
Pain duration		
1-3 months	%	7.4
3-6 months	%	18.9
0.5-1 year	%	23.4
1-2 year	%	19.1
2-5 year	%	14.9
More than 5 years	%	16.3
Pain average past week (0-10) [d]	Mean (sd)	5.4 (2.3)
Pain worse past week (0-10) [d]	Mean (sd)	6.8 (2.5)
PDI score (0-70) [e]		
Total sample		
Baseline	Mean (sd)	34.7 (11.7)
	Range	3-60
Discharge	Mean (sd)	24.2 (14.1)
Mean change [f]	Mean (sd)	-10.5 (13.8)*
	95% CI of mean change	9.1 - 12.0
Baseline PDI Q1		
Baseline	Mean (sd)	19.3 (6.2)
	Range	3-27
Discharge	Mean (sd)	16.4 (12.2)
Mean change	Mean (sd)	-2.9 (12.3)*
	95% CI of mean change	0.3 - 5.5
Baseline PDI Q2		
Baseline	Mean (sd)	32.0 (2.1)
	Range	28-35

Table 1. (Continued)

	Unit of measurement	Vocational rehabilitation (n=351)
Discharge	Mean (sd)	21.0 (11.8)
Mean change	Mean (sd)	-11.0 (11.7)*
	95% CI of mean change	8.4 - 13.5
Baseline PDI Q3		
Baseline	Mean (sd)	38.9 (2.1)
	Range	36-42
Discharge	Mean (sd)	28.0 (13.5)
Mean change	Mean (sd)	-10.9 (13.8)*
	95% CI of mean change	7.9 – 13.9
Baseline PDI Q4		
Baseline	Mean (sd)	48.8 (4.5)
	Range	43-60
Discharge	Mean (sd)	31.4 (13.6)
Mean change	Mean (sd)	-17.5 (13.4)*
	95% CI of mean change	14.6 – 20.3

SD, Standard Deviation; PDI, Pain Disability Index; IQR, Interquartile Range; Q, Quartile.
[a] Spine, low back, upper back, neck and/or shoulder pain; [b] Lower extremities, hip(s), upper leg(s), and/or ankle(s); [c] Upper extremities, arm(s), and/or hand(s) or finger(s); [d] 0=no pain, 10=worst possible pain; [e] 0= no disability, 70= maximum disability; [f] PDI discharge score – PDI baseline score.
* Significant change between baseline (T0) and discharge (T1) ($p < 0.05$).

Floor and ceiling effects

Floor and ceiling effects were absent in this study. The PDI total baseline score (min-max) was 3-60; 2.6% of the study sample had a total PDI baseline score <10 and 0.3% (1 person) of the study sample had a total PDI baseline score of 60.

Measurement error

The SEM was 1.2, the SDC for group level was 0.3 and the SDC for individuals was 3.4 (Table 2).

Interpretability

The SDC individual was smaller than the MIC in the total sample and in all PDI baseline quartile subgroups (Table 2). Of the total study sample, 70% improved at or above the SDC individual and 42% improved at or above the MIC (Table 4). Of the baseline quartile subgroups, 55-82% improved at or above the SDC individual and 40-46% improved at or above the MIC. Table 3 shows the PDI baseline score of various (sub) groups.

Table 2. Responsiveness parameters PDI

Parameter	Total sample	Baseline PDI Q1 (3-27)	Baseline PDI Q2 (28-35)	Baseline PDI Q3 (36-42)	Baseline PDI Q4 (43-60)
Improved (N)	124	34	32	24	34
Stable (N)	217	55	49	59	54
AUC (CI)	0.79 (0.74-0.84)	0.70 (0.59-0.81)	0.87 (0.79-0.95)	0.83 (0.73-0.93)	0.85 (0.77-0.93)
MIC	12.5	6.5	14.5	14.5	19.5
Sensitivity	0.68	0.68	0.81	0.71	0.79
Specificity	0.73	0.67	0.80	0.73	0.81
SEM	1.2	1.2	1.2	1.2	1.2
SDC individual	3.4	3.4	3.4	3.4	3.4
SDC group	0.3	0.3	0.3	0.3	0.3

PDI, Pain Disability Index; Q, Quartile; AUC, Area Under the Curve; CI, Confidence Interval; MIC, Minimal Important Change; SEM, Standard Error of Measurement; SDC, Smallest Detectable Change

Table 3. Reference values baseline PDI scores

Diagnosis	N	PDI score Mean (SD)	Source
Chronic musculoskeletal pain	351	34.7 (11.7)	Present study
General population	2510	6.8 (11.4)	Mewes 2009 [24]
Acute back pain	178	38.0 (15.9)	Soer 2013 [6]
Chronic back pain	242	34.6 (13.8)	Soer 2012 [8]
Chronic low back pain	425	36.5 (13.8)	Soer 2013 [6]
Chronic pain	4867	38.9 (13.3)	Köke 2017 [13]
Widespread pain	365	41.4 (10.9)	Soer 2013 [6]
Pain average past week (0–10)			
Patients with pain score 1–4	589	27.6 (13)	Köke 2017 [13]
Patients with pain score 5–6	1291	34.7 (11.5)	Köke 2017 [13]
Patients with pain score 7–10	2759	43.2 (12.2)	Köke 2017 [13]

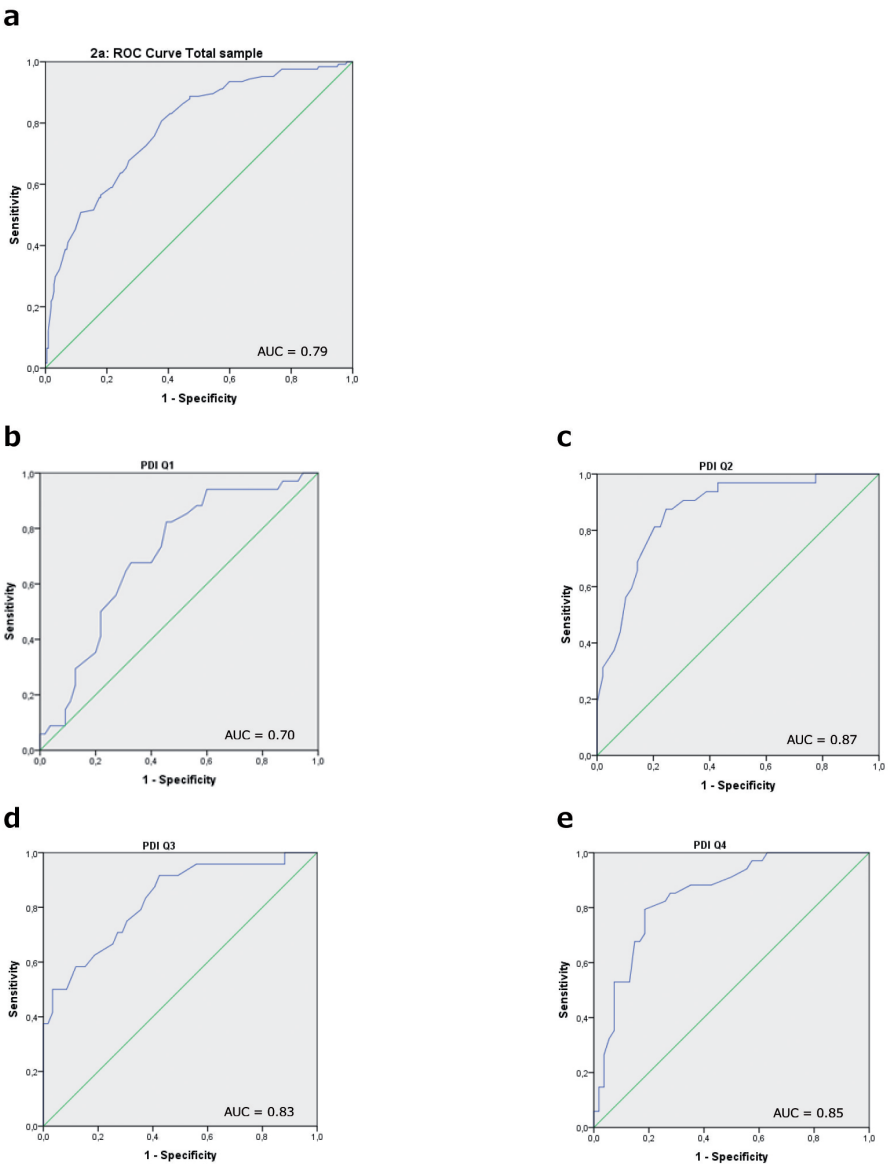
PDI, Pain Disability Index; SD, Standard Deviation

Table 4. Change scores in relation to MIC and SDC

	Total sample N=341	Baseline PDI Q1 (3-27) N=89	Baseline PDI Q2 (28-35) N=81	Baseline PDI Q3 (36-42) N=83	Baseline PDI Q4 (43-60) N=88
Change ≥1 point (%)	76.5	65.2	79.0	74.7	87.5
MIC	-12.5	-6.5	-14.5	-14.5	-19.5
≥MIC (%)	41.9	46.1	44.4	39.8	42.0
SDC individual	3.4	3.4	3.4	3.4	3.4
≥SDC individual (%)	69.8	55.1	74.1	68.7	81.8
MIC > SDC	Yes	Yes	Yes	Yes	Yes

PDI, Pain Disability Index; Q, Quartile; MIC, Minimal Important Change; SDC, Smallest Detectable Change

Figure 2. ROC curves of the PDI total sample and baseline quartiles



ROC, receiver operating characteristic; PDI, Pain Disability Index; Q, quartile; AUC, area under the curve. **a** ROC-curve of total study sample (n = 341). **b** ROC-curve of the sample with PDI baseline quartile 1 score (n = 89). **c** ROC-curve of the sample with PDI baseline quartile 2 score (n = 81). **d** ROC-curve of the sample with PDI baseline quartile 3 score (n = 83). **e** ROC-curve of the sample with PDI baseline quartile 4 score (n = 88)

Discussion

The results show that the PDI is responsive to detect clinically relevant changes in pain-related disability at discharge of vocational rehabilitation (AUC 0.79). A PDI change score of 13 points (MIC 12.5) can be considered as a real change in pain-related disability for the total study sample, and a PDI change score of 7-20 points can be considered as a real change in pain-related disability for PDI lowest and highest baseline quartile scores.

The responsiveness of the total study sample is in line with others [8] who found an AUC of 0.76 in patients with chronic back pain. However, the MIC of this study was 9.5 [8]. Because the sample size, external anchor's (both 7-item Likert scale), and PDI version (both Dutch language versions) were similar amongst both studies, we hypothesize that the difference in MIC might be caused by the difference in mean change score, namely 10.5 in the current study and 6.8 in the other study [8]. This difference in mean change score might be affected by the different sample characteristics, settings, and interventions, applied in the other study; VR on the one hand versus multidisciplinary rehabilitation, surgery, or anesthesiology [8]. Another explanation for the difference in MIC might be caused by the different ways in questioning the GPE anchor item, which was formulated in the current study as follows: "How are your (pain) complaints at this moment compared to pre-treatment?", and which was formulated in the other study as follows: "How much did your treated complaints change compared with pretreatment level?". Finally, the same data was collected in the present study between 2014 to 2017; despite the passage of time, the diversity of centers and professionals involved in the collection of data. These factors also could have influenced the findings on responsiveness. In summary, the different MIC and change scores between the present and discussed study show that the MIC and change score can differ per sample and setting.

The mean change score of the present study (10.5) is somewhat higher compared to a study that found a mean change score in PDI of 9.4 in patients with chronic pain after a multidisciplinary pain program [19]. This is surprising, because the study mentioned had a higher PDI baseline value, namely 37.8, which implicates more room for change, which we actually showed in the present study. Another study showed a mean change score in PDI of 14.0 (baseline score 47.6) in workers' compensation claimants with musculoskeletal disorders after a functional restoration program [25]. This PDI change score is slightly lower compared with the mean change score of 17.5 of the fourth quartile of the

present study, but it supports our finding that interpretation of the PDI change score is baseline dependent.

The interpretation of change score of the PDI can be interpreted as a "real" change in pain-related disability if the mean change score is at or above the MIC and if the SDC for individuals does not exceed the MIC (Figure 1, Table 2). It is difficult to compare our results with other studies, however, for two reasons. Firstly, we are only aware of one study that found an SDC of 17.9 in patients with acute back pain, chronic low back pain, and widespread pain [6]. The huge discrepancy compared with the current study (SDC 3.4) can be explained by the fact the study in question used the standard deviation of the mean PDI baseline score in the calculation of the SDC (personal communication with first author (RS)). We suppose that it is important for the calculation of the SDC to take the variability between time points into account [11]. Secondly, change scores of longitudinal cohort studies are regularly reported on group level (i.e. mean scores), whereas it is much more interesting to report the percentage of improved patients (according to the MIC), because this "… provides readers with values which are more easily understood and additional information to help them decide whether a treatment should be used." [22].

The baseline PDI score of the current study is similar compared to patients with chronic back pain [6, 8], but somewhat lower compared to patients with chronic pain and widespread pain. One reason for this difference might be a difference in patients executing paid work, which was 91% in the current study and 48% and 43% in chronic pain and widespread pain [6, 13]. Another difference might be due to a difference in pain baseline score of the present study compared with the chronic and widespread pain samples (5.4 versus 6.7 and 6.9, respectively). Köke et al. showed that higher pain score on baseline is related to significantly higher PDI baseline scores [13].

Methodological considerations

The first methodological consideration of this study was the assessment of the MIC. Two common methods can be used to calculate the MIC: the distribution-based method and the anchor-based method [20]. In the distribution-based method, 50% of the standard deviation of the baseline score (0.5*SD) of the measurement instrument serves as the MIC. In the anchor-based method an external anchor is used as the "gold standard" to discriminate between improved and unchanged persons, and the MIC can be obtained with an ROC

curve. Because the MIC can be derived from the sensitivity and specificity provided with an ROC curve, the MIC can be used in scientific research and clinical practice as a cut off point to determine the number of patients that have significantly changed. Patients with a change score greater than or equal to MIC can be called “responders”. With this method, the difference in percentages of responders between treatment groups can be determined [11]. Because of the aforementioned advantage, and because this method is recommended [20, 26-29], we used the anchor-based method in the present study. The second methodological consideration was how we dichotomized the anchor item into changed and unchanged groups, which we used for the calculation of the MICs. In the present study, the changed group consisted of patients who were “much improved” and “completely improved” and the unchanged group consisted of patients who were “little worsened”, “unchanged”, and “little improved”. Other papers, however, state that only a “little improved” group can serve as the (minimal important) change group [29, 30], or “little improved”, much improved” and “completely improved” as the changed group [20]. We, however, agree with Ostelo et al. who stated that “…”little improvement” is in the range of natural fluctuation, and that an “important” improvement should be greater than these (unimportant) fluctuations” [31: p. 92]. However, it is important to notice that the type of anchor-dichotomization directly influences the AUC and MIC. Therefore, the results of the present study must be interpreted with caution because the used cutoff has a high influence on the findings [20, 31]. The third and final methodological consideration was the number of baseline (sub) groups. We decided a priori to apply four subgroups (i.e. quartiles), because we had enough power. The number of four subgroups used in the present study was arbitrary, however. Nevertheless, there are no guidelines for conducting a particular number of (sub)groups based on baseline score, and there are as yet no subgroup scores known for the PDI based on pain-related disability (for example “low”, “intermediate” and “high” pain-related disability subgroups). Since the second and third baseline quartile of the present study showed similar MICs and mean change scores, future studies might propose to assess the responsiveness of three PDI baseline subgroups based on interquartile range (25^{th}, 50^{th}, and 75^{th} percentile).

Clinical message

Practitioners can use the following cutoff scores to decide if a PDI change score is clinically relevant at discharge of VR: patients with a baseline score of ≤27 should decrease minimal 7 points, patients with a baseline score between 28-42

should decrease minimal 15 points, and patients with a baseline score ≥43 should decrease minimal 20 points.

Conclusion

The PDI is a responsive questionnaire which can detect real change in decrease of pain-related disability in patients with CMP at discharge of vocational rehabilitation. Future research should focus on assessing the SDC and the MIC of the PDI in various patient samples and settings. Also, when using longitudinal cohorts, researchers are encouraged to report the portion of the sample with a change score at or above the MIC since this will enhance comparability and clinical relevance.

Abbreviations

AUC: Area under the curve; CMP: chronic musculoskeletal pain; COSMIN: COnsensus-based Standards for the selection of health Measurement INstruments; GPE: global perceived effect; ICC: intraclass correlation coefficient; MIC: Minimal Important Change; PDI: Pain Disability Index; QQ plot: quantile-quantile plot; ROC: receiver operating characteristics curve; SD: standard deviation; SDC: Smallest Detectable Change; SEM: Standard Error of Measurement; SPSS: Statistical Package for the Social Sciences; SStotal: sum of squares total; VR: vocational rehabilitation.

Ethics approval and consent to participate

Patients gave consent to participate in this study. Because this study contains routinely collected and anonymous data of care as usual programs, the Medical Ethical Committee of the Academic Medical Center, Amsterdam, the Netherlands, authorized this study and decided that a full application was not required (reference number: A1 17.405).

Consent for publication

Not applicable.

Availability of data and material

Please contact author for data requests.

Competing interests

The authors declare that they have no competing interests.

Funding

None.

Authors' contributions

TB initiated the study, conducted the statistical analyses and wrote the paper. All authors contributed to drafting the manuscript and have read and approved the final manuscript.

Acknowledgements

We thank the following vocational rehabilitation centers for their help in data collection: Libra (Eindhoven, Tilburg), Roessingh (Enschede), Adelante (Hoensbroek), Merem (Almere, Huizen), Zintens Groot Klimmendaal (Arnhem), UMCG Center for Rehabilitation (Groningen), and Heliomare (Wijk aan Zee).

References

1. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain. 2006 May;10(4):287-333.
2. Chapman JR, Norvell DC, Hermsmeyer JT, Bransford RJ, DeVine J, McGirt MJ, et al. Evaluating common outcomes for measuring treatment success for chronic low back pain. Spine (Phila Pa 1976). 2011 Oct 01;36(21 Suppl):S54-68.
3. Pollard CA. Preliminary validity study of the pain disability index. Percept Mot Skills. 1984 Dec;59(3):974.
4. Chibnall JT, Tait RC. The Pain Disability Index: factor structure and normative data. Arch Phys Med Rehabil. 1994 Oct;75(10):1082-6.
5. Jerome A, Gross RT. Pain disability index: construct and discriminant validity. Arch Phys Med Rehabil. 1991 Oct;72(11):920-2.
6. Soer R, Koke AJ, Vroomen PC, Stegeman P, Smeets RJ, Coppes MH, et al. Extensive validation of the pain disability index in 3 groups of patients with musculoskeletal pain. Spine (Phila Pa 1976). 2013 Apr 20;38(9):E562-8.
7. Gronblad M, Hupli M, Wennerstrand P, Jarvinen E, Lukinmaa A, Kouri JP, et al. Intercorrelation and test-retest reliability of the Pain Disability Index (PDI) and the Oswestry Disability Questionnaire (ODQ) and their correlation with pain intensity in low back pain patients. Clin J Pain. 1993 Sep;9(3):189-95.
8. Soer R, Reneman MF, Vroomen PC, Stegeman P, Coppes MH. Responsiveness and minimal clinically important change of the Pain Disability Index in patients with chronic back pain. Spine (Phila Pa 1976). 2012 Apr 15;37(8):711-5.
9. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010 Jul;63(7):737-45.
10. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007 Jan;60(1):34-42.
11. Terwee CB, Roorda LD, Knol DL, De Boer MR, De Vet HC. Linking measurement error to minimal important change of patient-reported outcomes. J Clin Epidemiol. 2009 Oct;62(10):1062-7.
12. Escorpizo R, Reneman MF, Ekholm J, Fritz J, Krupa T, Marnetoft SU, et al. A conceptual definition of vocational rehabilitation based on the ICF: building a shared global model. J Occup Rehabil. 2011 Jun;21(2):126-33.
13. Koke AJ, Smeets RJ, Schreurs KM, van Baalen B, de Haan P, Remerie SC, et al. Dutch Dataset Pain Rehabilitation in daily practice: Content, patient characteristics and reference data. Eur J Pain. 2017 Mar;21(3):434-44.
14. Waddel G, Burton AK: Is work good for your health and well-being? The Stationery Office 2006.
15. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010 May;19(4):539-49.

16. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol. 2010 Mar 18;10:22.
17. Beemster TT, van Velzen JM, van Bennekom CA, Frings-Dresen MH, Reneman MF. Cost-effectiveness of 40-hour versus 100-hour vocational rehabilitation on work participation for workers on sick leave due to subacute or chronic musculoskeletal pain: study protocol for a randomized controlled trial. Trials. 2015 Jul 28;16:317.
18. Reneman MF, Beemster TT, Edelaar MJ, van Velzen JM, van Bennekom C, Escorpizo R. Towards an ICF- and IMMPACT-based pain vocational rehabilitation core set in the Netherlands. J Occup Rehabil. 2013 Dec;23(4):576-84.
19. van Kampen DA, Willems WJ, van Beers LW, Castelein RM, Scholtes VA, Terwee CB. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). J Orthop Surg Res. 2013 Nov 14;8:40.
20. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. J Clin Epidemiol. 2010 May;63(5):524-34.
21. Baker DW, Hays RD, Brook RH. Understanding changes in health status. Is the floor phenomenon merely the last step of the staircase? Med Care. 1997 Jan;35(1):1-15.
22. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Ann Rheum Dis. 2005 Jan;64(1):29-33.
23. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005 Feb;19(1):231-40.
24. Mewes R, Rief W, Stenzel N, Glaesmer H, Martin A, Brahler E. What is “normal” disability? An investigation of disability in the general population. Pain. 2009 Mar;142(1-2):36-41.
25. Park J, Esmail S, Rayani F, Norris CM, Gross DP. Motivational Interviewing for Workers with Disabling Musculoskeletal Disorders: Results of a Cluster Randomized Control Trial. J Occup Rehabil. 2017 May 26.
26. de Vet HC, Ostelo RW, Terwee CB, van der Roer N, Knol DL, Beckerman H, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res. 2007 Feb;16(1):131-42.
27. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain. 2008 Feb;9(2):105-21.
28. King MT. A point of minimal important difference (MID): a critique of terminology and methods. Expert Rev Pharmacoecon Outcomes Res. 2011 Apr;11(2):171-84.
29. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008 Feb;61(2):102-9.
30. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989 Dec;10(4):407-15.

31. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine (Phila Pa 1976). 2008 Jan 01;33(1):90-4.

CHAPTER 5

Cost-effectiveness of 40-hour versus 100-hour Vocational Rehabilitation on Work Participation for Workers on Sick Leave due to Subacute or Chronic Musculoskeletal Pain: Study Protocol for a Randomized Controlled Trial

T. Beemster
J. van Velzen
C. van Bennekom
M. Frings-Dresen
M. Reneman

Trials. 2015 Jul 28;16:317

Abstract

Background

Although vocational rehabilitation is a widely advocated intervention for workers on sick leave due to subacute or chronic non-specific musculoskeletal pain, the optimal dosage of effective and cost-effective vocational rehabilitation remains unknown. The objective of this paper is to describe the design of a non-inferiority trial evaluating the effectiveness and cost-effectiveness of 40-hour multidisciplinary vocational rehabilitation compared with 100-hour multidisciplinary vocational rehabilitation on work participation for workers on sick leave due to subacute or chronic musculoskeletal pain.

Methods

A non-inferiority study design will be applied. The study population consists of workers who are on part-time or full-time sick leave due to subacute or chronic non-specific musculoskeletal pain. Two multidisciplinary vocational rehabilitation programmes following the bio-psychosocial approach will be evaluated in this study: 40-hour vocational rehabilitation and 100-hour vocational rehabilitation, both delivered over a maximum of 15 weeks. 100-hour vocational rehabilitation comprises five modules: work participation coordination, graded activity, cognitive behavioural therapy, group education, and relaxation. 40-hour vocational rehabilitation comprises work participation coordination and a well-reasoned choice from the other four modules. Four rehabilitation centres participate in this study, each delivering both interventions. Patients will be randomized into one of the interventions, stratified for the duration of sick leave (<6 weeks or ≥6 weeks) and type of sick leave (part-time or full-time). The primary outcome is work participation, measured by self-reported sick leave days, and will be assessed at baseline, mid-term, discharge, and at 2, 4, 6, 8, 10, and 12 months follow-up. Secondary outcomes are work ability, disability, quality of life, and physical functioning, and will be assessed at baseline, discharge, and at 6 and 12 months follow-up. Cost outcomes are absenteeism, presenteeism, health care usage, and travelling costs. Cost-effectiveness will be evaluated from the societal and employer perspectives.

Discussion
The results obtained from this study will be useful for vocational rehabilitation practice, and will provide stakeholders with relevant insights into two versions of vocational rehabilitation.

Trial registration
Dutch Trial Register identifier: NTR4362 (registered 17 March 2014).

Keywords
Vocational rehabilitation, subacute musculoskeletal pain, chronic musculoskeletal pain, work participation, cost-effectiveness, non-inferiority, randomized controlled trial, multi-centre

Background

Chronic musculoskeletal pain is a major health problem associated with decreased functioning and quality of life, sick leave, and increased direct and indirect medical costs [1-4]. The majority of the costs (48-88%) are attributed to indirect costs due to sick leave from work or productivity loss while at work [5, 6] . Chronic musculoskeletal pain arises when acute musculoskeletal pain does not disappear within six weeks, which occurs in 10-20% of the cases [7]. After a duration of six weeks, it is considered subacute musculoskeletal pain (SMP), and if the pain is still present after 12 weeks it is considered chronic musculoskeletal pain (CMP) [8]. If there is no clear medical explanation, the chronic musculoskeletal pain is called ''non-specific''.

Vocational rehabilitation is a widely advocated intervention for sick-listed workers with subacute or chronic non-specific musculoskeletal pain [9-12]. Vocational rehabilitation is ''a multi-professional evidence-based approach that is provided in different settings, services, and activities to working-age individuals with health-related impairments, limitations, or restrictions with work functioning, and whose primary aim is to optimize work participation'' [13]. In addition, work participation is conceptualized as the involvement in work roles or the lived experience of work. Work participation restriction refers to problems an individual may experience at work. Examples include number of hours lost from work (i.e. absenteeism), underperforming job expectations, reduced desired employment

(e.g. part-time employment, short-term disability, long-term disability, premature retirement, or fewer working hours than desired), and reduced career growth [14]. However, in this paper work participation (restriction) is expressed as the number of sick leave days due to subacute or chronic musculoskeletal pain. Research shows that vocational rehabilitation improves return to work [9-12, 15-22], and thus facilitates work participation. However, the dose-effect relation of vocational rehabilitation on work participation is unclear. Several reviews on the effectiveness of vocational rehabilitation on work participation for sick-listed workers with SMP and CMP show a wide range in treatment hours [9, 11, 20, 23]. In addition, a systematic review revealed a range of 6.4 to 196.8 hours in pain rehabilitation programmes [23]. So far, only one randomized controlled trial has compared the dose-effect relation of vocational rehabilitation (VR) [16]. Sick-listed workers with CMP were classified at baseline as good, medium, or poor based on their prognosis for return to work (i.e. return to work defined by the authors as absence of sick pay or related benefits in a given month), and were thereafter randomized to extensive VR (~120 treatment hours), light VR (~20-30 treatment hours), or care as usual (referred back to general practitioner). After 14 months follow-up, the participants classified with poor prognosis benefited most from the extensive VR, resulting in higher return to work rates, whereas patients classified with medium prognosis benefited from both the light and extensive programmes on improving return to work rates. In another paper, but using the same study construct and population as in the Haldorsen trial [16], results were conducted without the prognosis on return to work (i.e. good, medium, or poor), and on a follow-up period of two years. After two years follow-up, the light VR resulted in the highest return to work rates compared with usual care, but significance was only found in men. Additionally, the authors found no significant difference on return to work rates between light and extensive VR or between extensive VR and usual care [24].

As resources in health care are scarce, it is necessary to provide stakeholders information about the cost-effectiveness of intervention programmes. Economic evaluations (i.e. cost-effectiveness studies) provide information on the relative efficiency of two or more alternative interventions. The main aspects of any economic evaluation are to identify, measure, value, and compare the costs and consequences of alternatives [25]. A randomized controlled trial found that a participatory approach (~40 treatment hours consisting of a workplace intervention and graded activity) for sick-listed patients with chronic back pain was cost-effective on work participation (i.e. return to work) compared with usual care [18]. Similar interventions conducted in subacute low back pain

patients also show promising results on cost-effectiveness [21, 24, 26, 27]. However, there are no studies known that compare the cost-effectiveness of two (or more) vocational rehabilitation programmes. To provide relevant stakeholders (i.e. patients, referrers, employers, vocational rehabilitation centres, health care insurers, and policy makers) with information about effective and cost-effective vocational rehabilitation, comparison of two versions of vocational rehabilitation is needed.

Objectives

The objective of this paper is to describe the design of a multi-centre, randomized, non-inferiority study to evaluate the effectiveness and cost-effectiveness of 40-hour vocational rehabilitation compared with 100-hour vocational rehabilitation on work participation for patients with subacute or chronic musculoskeletal pain and with sick leave from work. We hypothesize that 40-hour VR will be non-inferior on work participation, and cost-effective in comparison with 100-hour VR.

The research questions are:

I) For workers on sick leave due to subacute or chronic musculoskeletal pain, is 40-hour vocational rehabilitation non-inferior on work participation compared with 100-hour vocational rehabilitation?

II) For workers on sick leave due to subacute or chronic musculoskeletal pain, is 40-hour vocational rehabilitation cost-effective compared with 100-hour vocational rehabilitation?

Methods

CONSORT

In the description of our study design, we follow the Consolidated Standards of Reporting Trials (CONSORT statement) with the extension of reporting on non-inferiority trials [28].

Organisation of the study

Approval for the study has been obtained by the Medical Ethics Committee of the Academic Medical Center, Amsterdam, the Netherlands (approval number: 2013_366). The trial is registered in the Dutch Trial Register (http://www.trialregister.nl/trialreg/index.asp) with identification number NTR4362. All participants will sign written informed consent forms and will be insured according to Dutch Law in case of any damage caused by participation in the study. Figure 1 shows a flow chart of the design of the study.

Study design

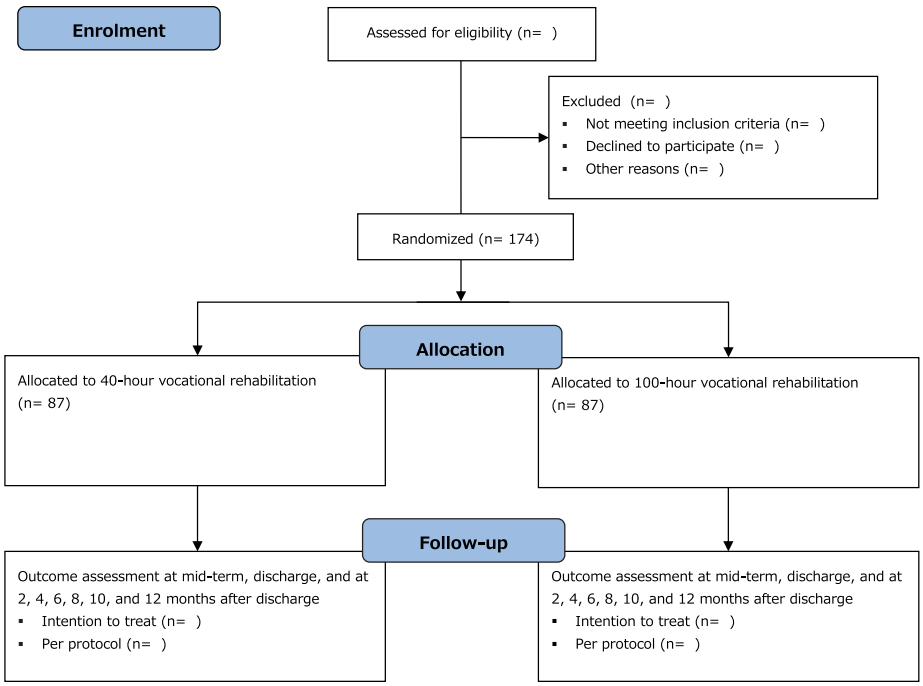
A multi-centre, randomized, 12-month follow-up, non-inferiority study design will be performed to evaluate the effectiveness and cost-effectiveness on work participation of 40-h versus 100-h vocational rehabilitation for patients with subacute or chronic musculoskeletal pain and on sick leave from work.

Study population

The inclusion criteria for this study are: 1) individuals of working age (18-65 years); 2) suffering from subacute (6-12 weeks) or chronic (>12 weeks) non-specific musculoskeletal pain such as back, neck, shoulder, widespread pain, Whiplash Associated Disorder (WAD I or II), or fibromyalgia; 3) having paid work (employed or self-employed) for at least 12 hours per week; 4) the expectation that the employment or self-employment will not be terminated in the year following the vocational rehabilitation programme; 5) having short-term (<6 weeks) or long-term (≥6 weeks) part-time or full-time sick leave; 6) being able to understand Dutch and able to complete questionnaires in Dutch; 7) having the motivation to participate in vocational rehabilitation aimed at optimizing work participation; 8) reimbursement of programme costs by the employer (i.e. the work participation coordination module, see Appendix 1); 9) having an

email address; and 10) having granted informed consent. The exclusion criteria for this study is having comorbidities that are the primary reason for sick leave, such as acute or specific medical problems, clinical depression or burnout, severe asthmatic symptoms, diagnosed chronic fatigue, and neuropathy.

Figure 1. Flow chart of the design of the study



Setting

Patients will be recruited between November 2014 and August 2016. The study will be performed in four vocational rehabilitation centres in the Netherlands that are part of a nationwide network of twelve VR centres. The four participating centres in this study are geographically spread across the Netherlands and have been selected according to the number of patients expected to be referred in 2014-2016.

Recruitment of participants

Recruitment of participants occurs in five steps; the first three steps are regular steps and the last two steps have been added especially for this study. Step 1. Patients will be referred to one of the four participating centres by either an occupational physician, medical specialist, general practitioner, or employer. Step 2. A rehabilitation physician (RP) will assess the patient's medical history, bio-psychosocial restrictions, and work-related limitations. Step 3. A multidisciplinary screening comprising a mental, physical and occupational assessment will take place, which will be performed by a psychologist, physiotherapist, and occupational specialist. Step 4. After completing the multidisciplinary screening, the patient will be provided with verbal and written information about the study. When all study criteria have been met – which will be decided by the RP – the patient will be asked to sign the informed consent form. Step 5. When the patient has granted written informed consent, the patient will be randomized into 40-h or 100-h vocational rehabilitation.

Interventions

40-hour vocational rehabilitation

40-hour vocational rehabilitation is a multidisciplinary bio-psychosocial [29] group-based programme, and consists of work participation coordination (10 hours), and a choice of 30 hours of a set of modules offered in the 100-h vocational rehabilitation programme, such as graded activity, cognitive behavioural therapy, group education, and relaxation. These modules are described in detail in Appendix 1. Since the choice of 30 hours of modules will be prioritized by the multidisciplinary screening team after the multidisciplinary screening at baseline, the content may differ per patient. 40-h VR lasts a maximum of 40 hours in 15 weeks. Each rehabilitation centre will prioritize the number of sessions per participant per week, but the following framework will be a guideline for the rehabilitation centres: week 1-5 two sessions/week, week 6-10 one session/week, week 11-15 2-3 sessions in five weeks. The 40-h VR programme will be extended if: a patient's percentage of working hours per week pertaining to contract hours at discharge compared with working hours per week pertaining to contract hours at baseline is extended by 25-50%, and the multidisciplinary team expresses strong arguments that the patient is likely to benefit from the extension. However, this protocol deviation should occur in no more than 5% of the cases. This percentage is arbitrarily chosen by the authors of this paper; if

more than 5% of the participants deviate from the study protocol, the robustness on non-inferiority (i.e. research question 1) will decline [30].

100-hour vocational rehabilitation

100-h vocational rehabilitation is a multidisciplinary bio-psychosocial group-based programme, and encompasses a set of modules: work participation coordination, graded activity, cognitive behavioural therapy, group education, and relaxation. These modules are described in detail in Appendix 1. 100-h VR consists of approximately 100 hours, and is an existing VR programme in the Netherlands conducted by twelve rehabilitation centres, four of which will participate in this study. 100-h VR is delivered over a period of 15 weeks with two sessions (~3.5 h/session) per week. 100-h VR appears similar to other VR trials [17, 31], but has a longer duration (in weeks) and consists of more graded activity hours as compared with similar studies [18, 21, 32].

Data collection

Self-reported data will be collected using web-based questionnaires. Data will be collected at baseline (i.e. before and during the multidisciplinary screening, T0), 7 weeks after the start of the intervention (mid-term, T1), 14 weeks after the start of the intervention (discharge, T2), and at 2, 4, 6, 8, 10 and 12 months follow-up after discharge (T3-T8). Figure 2 shows the timing of the data collection. In addition, pilot data shows an expected delay of ~1.5 months between baseline and the start of the intervention. At each data point, participants will receive an email with login data and the request to complete questionnaires on a website. If participants do not complete the questionnaire within a week, they will receive a reminder email. If the questionnaire is not completed after this reminder, patients will be telephoned by a researcher (TB), who will ask patients to complete the questionnaire. Table 1 presents the outcome measures of the data collection.

Figure 2. Timing of data collection

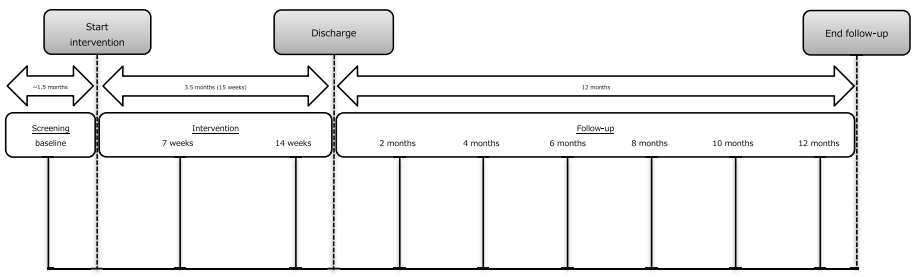


Table 1. Outcome measures for each of the measurement moments

Outcomes	Time measured			Follow-up					
	Baseline	Mid-term	Discharge	2 months	4 months	6 months	8 months	10 months	12 months
	T0	T1	T2	T3	T4	T5	T6	T7	T8
Descriptive variables									
Demographic variables	x								
General perceived health	x								
Work-related psychosocial factors	x								
Self-efficacy of work participation	x								
Pain intensity	x		x						
Fatigue	x								
Work tolerance functions	x								
Outcome measures									
Primary									
Work participation*	x	x	x	x	x	x	x	x	x
Secondary									
Work ability	x		x			x			x
Disability	x		x			x			x
Physical functioning	x		x			x			x
Quality of life	x		x			x			x
Costs*									
Presenteeism	x	x	x	x	x	x	x	x	x
Health care usage	x		x		x		x		x

*The primary outcome work participation will also serve as a cost outcome (i.e. absenteeism)

Outcome measures

The selection process of the questionnaires used in this study, and information about their validity and reliability is described in a core set paper [33]. The following outcomes will be assessed to answer the research questions:

Primary outcome

The primary outcome in this study is work participation expressed as total sick leave days due to subacute or chronic musculoskeletal pain. Total sick leave days will be calculated from the start of the intervention until 12 months follow-up after discharge. Sick leave will be measured using the absenteeism items of the iMTA (institute for Medical Technology Assessment) Productivity Cost Questionnaire (iPCQ) [34]. The questionnaire has a recall period of 4 weeks and measures sick leave on working days and on a generic basis (i.e. the reason for sick leave is not asked). We have made slight adaptations to measure sick leave specifically (i.e. related to subacute or chronic musculoskeletal pain, or other reasons such as flu), and we have added an item to assess the working hours at this moment: 'Are you working for the full number of hours you were contracted for?', with three possible answers: 'yes', 'no, I am partly at work', and 'no, I am on 100% sick leave'). After the answer 'no, I am partly at work' the participant is asked to fill in the number of hours they are working per week at that moment. The iPCQ is the result of combining two existing Dutch questionnaires (PRODISQ and SF-HLQ), and is recommended by the Dutch guideline for health economic evaluations [35]. The iPCQ has been translated by a professional language institution into a patient-friendly version using more simple language, thereby increasing the feasibility and validity of the questionnaire [34]. A population with mental health problems showed a satisfactory reliability regarding the iPCQ absenteeism items (icc 0.83) [36]. Until now, reliability of the iPCQ has not been tested in our study population. This needs to be done in the near future.

Secondary outcomes

Work ability will be measured using a single item of the Work Ability Index (WAI) [37]. The current work ability compared to lifetime best work ability can be scored on a 0-10 response scale, where 0 represents 'completely unable to work' and 10 represents 'work ability at its best'.

Disability will be measured using the Pain Disability Index (PDI). The PDI is a 7-item questionnaire for investigating the magnitude of the self-reported pain-related disability, independent of region of pain or pain-related diagnosis. The PDI measures family/home responsibilities, recreation, social activity, occupation, sexual behaviour, self-care, and life support activity. The questionnaire is constructed according to a 0-10 numeric rating scale in which 0 means no disability and 10 maximum disability. Total scores can range from 0 to 70, with higher scores reflecting higher interference of pain with daily activities [38, 39].

Physical functioning will be measured using the physical functioning subscale of the RAND-36. The questionnaire assesses self-reported physical functioning independent of (pain) diagnosis [40]. The physical functioning scale consists of 10 questions with three possible answers: 'yes, limited a lot', 'yes, limited a little', and 'no, not limited at all'. The total score can range from 0 to 100, with higher scores indicating better physical functioning. The validity and reliability of the Dutch version are good [41].

Quality of life will be measured using the validated Dutch version of the EuroQol-5D (EQ-5D) [42, 43]. The EQ-5D measures five dimensions: mobility, self-care, activities of daily life, pain and anxiety/depression on a categorical scale (1-3). The EQ-5D is a widely employed instrument used to assess health-related quality of life (QoL), and is recommended by the Dutch guideline for health economic evaluations [35]. To allow comparison between several conditions and interventions, Quality Adjusted Life Years (QALYs) will be calculated in three steps. First, the EQ-5D scores (measured at baseline, discharge, 6 months follow-up, and 12 months follow-up) will be converted to utility scores using the Dutch EQ-5D tariff [43]. Second, QALYs will be calculated from three time periods (1 = baseline - discharge, 2 = discharge - 6 months follow-up, 3 = 6 months follow-up – 12 months follow-up). Third, one summated QALY will be calculated from the calculated QALYs in step two.

Cost outcomes

The following outcomes will be assessed to evaluate the cost-effectiveness of 40-h VR compared with 100-h VR.

Absenteeism data will be derived from the work participation (i.e. primary outcome) data in this study.

Presenteeism will be assessed using the presenteeism items of the iPCQ [34]. The questionnaire measures the total days of mental or physical complaints at work, with a recall period of 4 weeks. The amount of work performed accompanied by mental or physical complaints is measured on a 0-10 response scale, where 0 represents 'I couldn't do anything', to 5 'I could do about half as normal', to 10 'I could do the same as normal'. A population with mental health problems showed good feasibility and validity [34], and moderate reliability for the number of days while impeded by mental or physical complaints (icc 0.56), and a satisfactory reliability for the efficiency rate (0-10) item (icc 0.73) [36].

Health care usage will be assessed using the Trimbos iMTA questionnaire for measuring Costs of Psychiatric Illnesses (TiC-P), module 1 [36, 44]. A recall period of 4 weeks is used in this questionnaire. Visits and consultations of the following health care providers were measured: general practitioner, physiotherapist, manual therapist, exercise therapist, occupational therapist, psychologist, insurance physician, medical specialists in hospitals, hospitalization (number of days), occupational physician, social worker, and dietician. Additional items were alternative care, home care, medication use, and job-related care like job coaches, ergonomic changes at the work site and reintegration specialists. Slight adaptations in the context and scope of health care practitioners were made to better match TiC-P to the target population (i.e. from psychiatry to pain and work). Another modification was that visits and consultations were measured in both generic and sickness-specific terms. Research shows that health care usage assessment by means of self-reported questionnaires is reliable [45]. A population with mental health problems showed good feasibility, promising construct validity, good agreement on medical resource use (yes/no), and sufficient test-retest reliability on the number of contacts with the health care providers [36]. Until now, reliability of the TiC-P has not been tested in our study population. This needs to be done in the near future.

Patient characteristics

Patient characteristics will be collected at baseline (i.e. before and during the multidisciplinary screening) to evaluate if randomization resulted in two prognostically comparable groups. The following characteristics will be collected:

Demographic variables: age, gender, marital status, nationality, body mass index (obtained from self-reported weight and height), educational level, and health condition [46, 47].

General perceived health will be assessed using a single item of the RAND-36 questionnaire [41].

Work-related psychosocial factors will be assessed using the Work Reintegration Questionnaire (WRQ). The questionnaire consists of 78 items distributed across eight scales: distress, illness behaviour/coping, job strain, job satisfaction, job control, avoidance, perfectionism, and stressful home situation. The questionnaire was developed and validated in Dutch (VAR: vragenlijst arbeidsreintegratie) [48, 49].

Self-efficacy of work participation will be assessed on a 0-10 response scale. Participants rate the certainty that they will be working in six months time, where 0 represents 'not at all certain' to 10 'extremely certain'. A score of ≥5 is associated with successful work participation after six months for workers with subacute back pain [50].

Pain intensity and fatigue will each be measured using two questions from an 11-point Numeric Rating Scale, ranging from 0 'no pain/fatigue' to 10 'worst possible pain/fatigue', requiring patients to rate their worst and average intensity of the last seven days [51].

Work tolerance functions will be assessed at baseline during the multidisciplinary screening using standardized lifting capacity tests from the Functional Capacity Evaluation (FCE) test battery: lifting low and/or overhead lifting. The lifting tests to be assessed depend on the individuals' work tasks. Procedures are described in detail elsewhere [52]. Lifting tests were found to be predictive of work participation in patients with musculoskeletal disorders [53].

Non-inferiority hypothesis

A reduction in sick leave days of more than 30 days per year is deemed a clinically significant improvement on work participation [15, 17-19, 24]. A difference in sick leave days of 30 or less (from the start of the intervention until 12-months follow-up) between 40-h and 100-h VR is assigned as the margin of non-inferiority in this study. Our hypotheses are:

H0: $\mu 1 - \mu 2 \geq 30$
H1: $\mu 1 - \mu 2 < 30$

H0 is the null hypothesis, and H1 is the alternative hypothesis, μ1 is the mean number of sick leave days in the 40-h VR group, and μ2 is the mean number of sick leave days in the 100-h VR group.

Non-inferiority is claimed if the upper bound of the one-sided 95% confidence interval of the treatment effect difference (μ1 - μ2) on work participation does not exceed 30, which means that the risk of it being inferior is within acceptable boundaries [30]. We expect a normal distribution of the primary outcome work participation. If the data on the primary outcome does not follow a normal distribution we will perform log transformations. The margin of non-inferiority will then be interpreted as a 28% increase in sick leave days difference of μ1 - μ2. We calculated this percentage as follows:

30 / 107 = 28%. Where 30 denotes the margin of non-inferiority and 107 the expected mean days of sick leave in the 100-h VR arm [15, 17-19, 24] during the timing of the data collection, which equals ~15.5 months (intervention period of 3.5 months + follow-up period of 12 months, see figure 2).

Statistical methods

All statistical analysis will be performed at the patient level, with descriptive statistics being used to compare the baseline measurements of the two intervention groups. If necessary, analyses will be adjusted for baseline differences. All analysis will be performed according to the intension-to-treat principle and the per protocol principle [30, 54]. To claim non-inferiority, both intention-to-treat and per protocol analysis must show non-inferiority [54]. Missing data on costs and effects will be assessed using multiple imputation techniques [25]. The imputation technique will depend on the results (i.e. missing completely at random, missing at random, or missing not at random).

Effectiveness

The primary outcome work participation will be analysed in three steps. Step 1. For every time point (i.e. T0 - T8, see figure 2) we will present the number of sick leave days as an absolute number and as a percentage related to contract hours/ month, in which the absolute number and percentages between a given time point and the preceding time point will be calculated using linear extrapolation, as recommended [36, 44, 55]. Step 2. We will calculate and present the cumulative total days of sick leave per month from the start of the intervention until 12 months follow-up using an area under the curve for all measurement points,

in which the number of sick leave days between a given time point and the preceding time point will be calculated using linear extrapolation. Step 3. Linear mixed models with multilevel analyses will be performed to assess non-inferiority between the two groups at 12 months follow-up (i.e. intervention period and 12 months follow-up) by means of 95% confidence intervals (i.e. CI approach). To improve generalizability and comparability of this study with other studies, we will repeat step 3 at the following time intervals: I) discharge - 12 months follow-up; II) start of intervention - 6 months follow-up; III) discharge - 6 months follow-up. These additional analyses will contain no conclusions about non-inferiority and will be analysed in the 'classical' superiority manner.

A t-test or Mann-Whitney U test (in the case of no normal distribution) will be used to examine differences at discharge, at six months follow-up, and at 12 months follow-up (defined as the difference in outcome between baseline and last follow-up) in all secondary outcomes between the intervention groups. We will perform these analyses on superiority, thus without margins of non-inferiority, as this is only relevant for the analysis of the primary outcome.

Cost-effectiveness

It is recommended to conduct various kinds of economic evaluations within the same study to inform all relevant stakeholders [25]. We will perform three types of cost analysis: cost-effectiveness analysis, cost-utility analysis, and cost-benefit analysis.

Cost-effectiveness analysis: societal perspective. The cost-effectiveness analysis (CEA) in this study will be evaluated from the societal perspective (i.e. all costs related to the intervention will be taken into account irrespective of who pays for them). Costs consist of direct medical costs (i.e. intervention costs, health care usage, and travelling costs) and indirect costs (i.e. productivity loss in paid work due to absenteeism and presenteeism). All costs will be summated for each individual patient. All summated costs will be indexed in euros for the reference year 2015. We will follow the friction cost method with a friction cost period of 160 days and an elasticity of 0.8 for the calculation of absenteeism costs [56], as recommended by the Dutch guideline for health economic evaluations [35], and described in detail elsewhere [34, 56]. To calculate the presenteeism costs, the costs of productivity losses will be multiplied by the number of workdays lost, with age and gender-specific productivity levels per paid employee indexed for the year 2015 [34, 35].

Both the incremental costs and incremental effects will be used to calculate the incremental cost-effectiveness ratio (ICER). The ICER will be calculated as (C1 − C0)/(E1 − E0), where C denotes the average per-participant costs and E denotes the effect on work participation in the 40-h and 100-h VR groups (subscripted 1 and 0). As absenteeism data will be used for the assessment of the effect ratio of the ICER, it will be excluded from the cost ratio part of the ICER. The ICER can be interpreted as the net costs (or savings) per extra unit of effect. In our study the extra unit of effect equals 1 day increase in work participation. To estimate uncertainty in the cost and effect data, non-parametric bootstraps will be used to simulate 5,000 ICERs [57]. To show statistical uncertainty on the results of cost-effectiveness, each simulated ICER will be plotted on a cost-effectiveness plane [58]. Although cost-effectiveness planes give a good impression of the uncertainty surrounding the ICER, they do not provide a summary measure of the joint uncertainty of costs and effects [25]. We will therefore perform cost-effectiveness acceptability curves (CEAC), which will provide insight into the probability that 40-h VR is cost-effective in comparison with 100-h VR [25].

Cost-utility analysis. A cost-utility analysis (CUA) will be conducted in which the incremental costs per QALY will be estimated and which will also be presented on a cost-effectiveness plane and CEAC. Public policymakers may be interested in CUA, because they can compare the results between several conditions and interventions [25].

Cost-benefit analysis: employer's perspective. As employers reimburse the work participation coordination module in both 40-h VR and 100-h VR, analysis from the employer's perspective (i.e. only the costs relevant to the employer will be considered, including intervention, absenteeism, and presenteeism costs) is useful. It is recommended to conduct cost-benefit analysis (CBA), in which both costs and consequences are measured in monetary units. In accordance with van Dongen et al. [25], we will perform return on investment (ROI) analyses, in which three ROI metrics are calculated; (1) net benefits (NB), (2) benefit-cost ratio (BCR), and (3) ROI.

NB = benefits − costs

BCR = benefits/costs

ROI = (benefits − costs)/costs [*100]

Costs will be defined as intervention costs. Benefits will be defined as the difference in monetized outcome measures (i.e. absenteeism and presenteeism costs) between 100-h and 40-h VR during the measurement period (i.e. intervention period and follow-up, see figure 2), with positive benefits indicating

reduced spending in the 40-h group. To estimate uncertainty, 95% CIs around the benefit estimates and NB will be estimated by means of bootstrap confidence intervals. Financial returns of 40-h VR are positive if the following criteria are met: NB > 0, BCR > 1, and ROI > 0% [25].

Sensitivity analyses
To assess the robustness of the results on cost-effectiveness, we will perform four sensitivity analyses. First, analyses will be performed using the complete cases only. Second, analyses will be performed in which the lost productivity costs will be calculated according to the human capital approach. In the human capital approach, total sick leave days are not fixed as in the friction cost approach, and elasticity is not required [25]. Third, analyses will be performed with sick leave and health care usage data that are related to subacute or chronic musculoskeletal pain. Fourth, the observed outliers with very high lost productivity will be excluded from the analysis.

Sample size

A sample size of 174 is calculated to be sufficient (with a one-sided 95% CI, 80% power, alpha of 0.025, standard deviation of 80 and a margin of non-inferiority of 30 days) to establish non-inferiority of 40-h VR. The sample size calculation allowed for 15% loss to follow-up – 10% expected from comparable studies [18, 59, 60] and 5% expected due to the extension of the programme in the 40-h VR group. An intraclass correlation coefficient (ICC) of 0.05 is accounted for by the use of four rehabilitation centres with two clusters (40-h and 100-h VR) at each centre [61, 62]. Because of the difference in programme hours between 40-h and 100-h VR, we expect 40-h VR to benefit by 8 extra working days available during the intervention period. We accounted for this in the power calculation by using minus 8 as the expected mean difference between 40-h and 100-h VR. In our power calculation we assumed a normal distribution of the primary outcome work participation. If the data on the primary outcome does not follow a normal distribution, we will perform log transformations. As previously stated, we will allow 28% as the margin of non-inferiority when the data is log transformed.

According to the number of patients expected to be referred to the four participating rehabilitation centres per year (approximately 350), and after accounting for two-thirds of non-participation in the study according to Lasagna Law [63], we expect an inclusion of 115 participants per year for this study.

Hence, our inclusion period will cover approximately 18 months, and the data collection period will cover 2 years and 9 months.

Randomization

An independent statistician prepared the randomization by using computer-generated randomization tables. To prevent unequal randomization, employees are pre-stratified by duration of sick leave (short-term <6 weeks or long-term ≥6 weeks) and whether they are on full-time (100%) or part-time (≤99%) sick leave. Block randomization with blocks of four will be applied to ensure equal group sizes within each stratum. A separate block randomization table is generated for each of the four participating vocational rehabilitation centres. For each stratum, the researcher will prepare opaque, sequentially numbered, and sealed coded envelopes, with a note for either 40-h VR or 100-h VR. After the multidisciplinary screening (at baseline), the multidisciplinary screening team and rehabilitation physician will fulfil all study criteria. If participants meet all criteria, they will be allocated to 40-h or 100-h VR. Treatment allocation will be performed by a member of the multidisciplinary screening team (MST) at each centre, and can be performed at the centre or via telephone (i.e. this will differ per centre). The MST member hands over two envelopes (left over) of that stratum, and the patient is asked to pick one of the envelopes, open the envelope and sign the note. In the case of telephone allocation, the MST member will ask the patient to sign informed consent and to return it via a reply envelope. When the signed informed consent is received, the MST member will perform the treatment allocation without the patient. After randomization, a research assistant will make an appointment for the patient's first intervention date.

Blinding

Blinding in this study is not possible because of the nature of the intervention. However, the data analyst will be kept blinded to the allocation. Participants will complete self-reported web-based questionnaires outside the study setting, so the multidisciplinary intervention team has no influence on the outcome assessment. After randomization, all participants are labelled with a research code consisting of a unique consecutive number. An independent researcher will maintain the coding scheme. Data analysis will be performed using this research code to guarantee that analyses of the data by the researcher will be blinded.

Co-interventions and compliance

The patients' self-reported health care usage data will be used for the assessment of co-interventions. Compliance will be assessed using information about attendance to the programme and compliance to the treatment protocol, and will be assessed after each intervention session in an electronic log by a member of the multidisciplinary intervention team (MIT). Furthermore, the MIT member will assess at discharge if the programme was completed as planned. This will be assessed on a binary scale: 'programme completed as planned', or 'programme deviated'. In the latter case, a closed question follows: programme deviated due to 'early discontinuation due to adverse events such as accident, surgery, or major private event', 'early discontinuation due to goals being achieved', 'extension of intervention programme due to non-achievement of goals', or 'other reasons'. In the case of an early discontinuation or extension of the programme, the number of deviated weeks will be reported. The information about compliance will be applied to perform the per protocol analyses.

Discussion

The purpose of the presented study is to evaluate the effectiveness and cost-effectiveness of 40-hour vocational rehabilitation versus 100-hour vocational rehabilitation on work participation for sick-listed workers due to subacute or chronic non-specific musculoskeletal pain. We hypothesize that there is non-inferiority on work participation after a 12-month follow-up period (including intervention period and 12 months follow-up) between both programmes, and we expect cost-effectiveness of 40-h VR in comparison with 100-h VR.

Context of this study

In the Netherlands, both employer and employee are responsible for the work participation process of the sick-listed employee during the first two years of sick leave. The employer and employee can be supported by a certified reintegration company and/or an occupational physician (OP). In the first two years of sickness, the employer is responsible for the costs of wage replacement, which is regulated by the Dutch Gatekeeper Improvement Act [64]. As a result of this act, the employer has to reimburse the work participation coordination module (costs: ~€1,200) for both interventions performed in this study. The other intervention modules are reimbursed by healthcare insurers.

Methodological considerations

The first methodological consideration of this study is that we were not able to fulfil the recommended steps for the composition of a margin of non-inferiority [30, 65]. This was because there is currently no historical data, such as meta analysis, comparing 100-h vocational rehabilitation with usual care. However, our non-inferiority margin is based on results from five randomized controlled trials evaluating multidisciplinary vocational rehabilitation compared with control interventions (i.e. usual care, such as occupational physician, physical therapist, occupational therapist, etc.) [15, 17-19, 24]. These studies found 43 days as the mean difference ((41.9 + 53.7 + 42 + 60.5 + 17.5)/5) in days on sick leave after one year follow-up in favour of multidisciplinary vocational rehabilitation. Consequently, we have decided that 30 is an acceptable margin of non-inferiority. One can argue that this limit is too wide, and that claiming non-inferiority could be achieved too simply, but for claiming non-inferiority the upper bound of the one-sided 95% confidence interval of the treatment effect difference (40-h VR – 100-h VR) must be 30 or less. Furthermore, when the margin of non-inferiority

of 30 is reached after 12 months follow-up, this is in fact 13 days (43-30) better than if a patient had been referred to usual care (occupational physician, physical therapist, occupational therapist, etc.). A mean saving of 10 sick days per year is considered the smallest effect that would be clinically worthwhile [66]. We therefore consider 30 as a reasonable margin of non-inferiority. The second methodological consideration of this study is the slight differences between the participating rehabilitation centres. For instance, each centre has its own logistic restrictions, such as a restriction in intervention facilities (i.e. equipment, building); and centres have evolved their own methods over the years. This may lead to interpretation issues in analysing the blended results of the four centres. However, we solved this problem by multilevel analyses and by performing both interventions at each centre, i.e. randomizing at the participant level. Although performing both interventions at the centre level may introduce contamination, we consider that the advantages of randomization at the participant level outweigh the disadvantages of contamination.

Strengths and limitations of this study

The first strength of this study is the assessment of the primary outcome work participation with self-reported questionnaires with a recall period of one month. This recall period will prevent recall bias. In addition, self-reported data about work participation has been shown to be a reliable alternative compared with electronic databases [67]. The second strength of this study is the analysis of the cost-effectiveness from both the societal and the employer's perspective. It is important to provide employers with information on the return on investment of both interventions, as this will help them to consider the right treatment. A third strength of this study is that we take presenteeism into account in the cost-effectiveness evaluation. Although most cost-effectiveness studies do not assess presenteeism [17, 18, 31, 35, 68], it is meaningful to take into account since the costs related to it are enormous, as shown by Lötters et al. [69], who found that for workers who returned to work after musculoskeletal disorders, the median loss for an 8-hour workday was 1.6 hours, and this remained at 12 months follow-up. A final strength of this study is the participation of four rehabilitation centres, all working with the bio-psychosocial model as a blueprint. This will increase the generalizability of this study.

A limitation of this study is that it is not possible to blind the patients and the multidisciplinary intervention and screening team. This may result in non-compliance to the treatment protocol, because patients may be aware of which

intervention parts they do not receive (especially in the 40-h VR group), whereas other patients in the same group will receive all intervention modules (see Appendix 1). When this deviation occurs on a large scale in the 40-h VR group, this will harm conclusions about non-inferiority. Another limitation of this study is that we do not correct for compensation costs, i.e. when colleagues take over the work of the less productive employee in their regular working hours. This may overestimate presenteeism costs [70].

Implications for practice

This study will provide essential knowledge about the dose-effect relation of vocational rehabilitation on work participation for workers on sick leave due to subacute or chronic musculoskeletal pain. The insights obtained from this study can be implemented in vocational rehabilitation practice, where centres would be able to judge which programme (40-h or 100-h) fit their patient groups best. Moreover, if our hypothesis about the effectiveness and cost-effectiveness of 40-h VR compared with 100-h VR is valid, this will be beneficial for patients, employers, and health care insurers. Patients will benefit from a decline in intervention hours, which will result in more time for work participation and leisure. Employers will benefit from a higher return on their investments, and health care insurers will benefit from higher volumes of patients who can participate in vocational rehabilitation within the same amount of time and money, or the same number of patients with lower costs.

Trial status

Participant enrolment started in November 2014. Recruitment is expected to be completed by the end of August 2016, and the trial will conclude by the end of December 2017.

Additional information

To place the results from the described cost-effectiveness study in perspective, the authors of this paper will also conduct a qualitative paper in which interviews with a random selection of the study population of the proposed RCT will be performed. The aim of these interviews will be to determine barriers and facilitators of 40-h and 100-h VR on work participation. The authors will also conduct focus group interviews with the multidisciplinary intervention teams to explore their experiences with both programmes.

Abbreviations

BCR, Benefit-cost ratio; CBA, Cost-benefit analysis; CBT, Cognitive behavioural therapy; CEA, Cost-effectiveness analysis; CEAC, Cost-effectiveness acceptability curves; CI, Confidence interval; CMP, Chronic musculoskeletal pain; CUA, Cost-utility analysis; EQ-5D, EuroQol-5D; FCE, Functional Capacity Evaluation; ICC, Intraclass correlation coefficient; ICER, Incremental cost-effectiveness ratio; iPCQ, iMTA (institute for Medical Technology Assessment) Productivity Cost Questionnaire; MIT, Multidisciplinary intervention team; MST, Multidisciplinary screening team; NB, Net benefits; OP, occupational physician; PDI, Pain Disability Index; QALYs, Quality Adjusted Life Years; QoL, Quality of life; ROI, Return on investment; RP, Rehabilitation physician; RTW, Return to work; SMP, Subacute musculoskeletal pain; TiC-P, Trimbos iMTA questionnaire for measuring Costs of Psychiatric Illnesses; WAD, Whiplash Associated Disorder; WAI, Work Ability Index; WRQ, Work Reintegration Questionnaire

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

TB, JVV, CVB, MFD, MR conceived the study, designed the study protocol, and drafted the manuscript. TB wrote the manuscript. JVV, CVB, MFD, and MR revised study protocols and wrote several sections of the manuscript. TB is in charge of coordination and direct implementation. JVV will help with data collection. All authors contributed to drafting the manuscript and have read and approved the final manuscript.

Acknowledgements

This study is financed by the Dutch rehabilitation centre Heliomare (Wijk aan Zee). The authors would like to thank Marcel Dijkgraaf and Nan van Geloven for their help with methodological issues. We thank Kumar Jamdagni for his check on grammar and spelling. We also wish to thank Tessa Lenderink, Monique Hodiamont-Joosten, Tanja Schiffelers, Marlique Langkamp, Levijn Romp, Margaret Lam, and Michel Edelaar for their help in streamlining the process in their vocational rehabilitation centres.

References

1. Bekkering GE, Bala MM, Reid K, Kellen E, Harker J, Riemsma R, et al. Epidemiology of chronic pain and its treatment in The Netherlands. Neth J Med. 2011 Mar;69(3):141-53.
2. de Vries HJ, Reneman MF, Groothoff JW, Geertzen JH, Brouwer S. Factors promoting staying at work in people with chronic nonspecific musculoskeletal pain: a systematic review. Disabil Rehabil. 2012;34(6):443-58.
3. Selander J, Marnetoft SU, Asell M. Predictors for successful vocational rehabilitation for clients with back pain problems. Disabil Rehabil. 2007 Feb 15;29(3):215-20.
4. Truchon M, Schmouth ME, Cote D, Fillion L, Rossignol M, Durand MJ. Absenteeism screening questionnaire (ASQ): a new tool for predicting long-term absenteeism among workers with low back pain. J Occup Rehabil. 2012 Mar;22(1):27-50.
5. de Vroome EM, Uegaki K, van der Ploeg CP, Treutlein DB, Steenbeek R, de Weerd M, et al. Burden of Sickness Absence Due to Chronic Disease in the Dutch Workforce from 2007 to 2011. J Occup Rehabil. 2015 Mar 25.
6. Lambeek LC, van Tulder MW, Swinkels IC, Koppes LL, Anema JR, van Mechelen W. The trend in total cost of back pain in The Netherlands in the period 2002 to 2007. Spine (Phila Pa 1976). 2011 Jun;36(13):1050-8.
7. Waddell G. 1987 Volvo award in clinical sciences. A new clinical model for the treatment of low-back pain. Spine (Phila Pa 1976). 1987 Sep;12(7):632-44.
8. Lambeek LC, Anema JR, van Royen BJ, Buijs PC, Wuisman PI, van Tulder MW, et al. Multidisciplinary outpatient care program for patients with chronic low back pain: design of a randomized controlled trial and cost-effectiveness study [ISRCTN28478651]. BMC Public Health. 2007;7:254.
9. Hoefsmit N, Houkes I, Nijhuis FJ. Intervention characteristics that facilitate return to work after sickness absence: a systematic literature review. J Occup Rehabil. 2012 Dec;22(4):462-77.
10. Kamper SJ, Apeldoorn AT, Chiarotto A, Smeets RJ, Ostelo RW, Guzman J, et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain. Cochrane Database Syst Rev. 2014;9:CD000963.
11. Meijer EM, Sluiter JK, Frings-Dresen MH. Evaluation of effective return-to-work treatment programs for sick-listed patients with non-specific musculoskeletal complaints: a systematic review. Int Arch Occup Environ Health. 2005 Aug;78(7):523-32.
12. Petit A, Fouquet N, Roquelaure Y. Chronic low back pain, chronic disability at work, chronic management issues. Scand J Work Environ Health. 2015 Mar;41(2):107-10.
13. Escorpizo R, Reneman MF, Ekholm J, Fritz J, Krupa T, Marnetoft SU, et al. A conceptual definition of vocational rehabilitation based on the ICF: building a shared global model. J Occup Rehabil. 2011 Jun;21(2):126-33.
14. AlHeresh RA, Keysor JJ. The Work Activity and Participation Outcomes Framework: a new look at work disability outcomes through the lens of the ICF. Int J Rehabil Res. 2015 Jun;38(2):107-12.

15. Bultmann U, Sherson D, Olsen J, Hansen CL, Lund T, Kilsgaard J. Coordinated and tailored work rehabilitation: a randomized controlled trial with economic evaluation undertaken with workers on sick leave due to musculoskeletal disorders. J Occup Rehabil. 2009 Mar;19(1):81-93.
16. Haldorsen EM, Grasdal AL, Skouen JS, Risa AE, Kronholm K, Ursin H. Is there a right treatment for a particular patient group? Comparison of ordinary treatment, light multidisciplinary treatment, and extensive multidisciplinary treatment for long-term sick-listed employees with musculoskeletal pain. Pain. 2002 Jan;95(1-2):49-63.
17. Jensen IB, Busch H, Bodin L, Hagberg J, Nygren A, Bergstrom G. Cost effectiveness of two rehabilitation programmes for neck and back pain patients: A seven year follow-up. Pain. 2009 Apr;142(3):202-8.
18. Lambeek LC, Bosmans JE, Van Royen BJ, Van Tulder MW, Van Mechelen W, Anema JR. Effect of integrated care for sick listed patients with chronic low back pain: economic evaluation alongside a randomised controlled trial. BMJ. 2010;341:c6414.
19. Loisel P, Abenhaim L, Durand P, Esdaile JM, Suissa S, Gosselin L, et al. A population-based, randomized clinical trial on back pain management. Spine (Phila Pa 1976). 1997 Dec 15;22(24):2911-8.
20. Palmer KT, Harris EC, Linaker C, Barker M, Lawrence W, Cooper C, et al. Effectiveness of community- and workplace-based interventions to manage musculoskeletal-related sickness absence and job loss: a systematic review. Rheumatology (Oxford). 2012 Feb;51(2):230-42.
21. Steenstra IA, Anema JR, van Tulder MW, Bongers PM, de Vet HC, van Mechelen W. Economic evaluation of a multi-stage return to work program for workers on sick-leave due to low back pain. J Occup Rehabil. 2006 Dec;16(4):557-78.
22. Waddell G, Burton AK. Is work good for your health and well-being? The Stationery Office 2006.
23. Waterschoot FP, Dijkstra PU, Hollak N, de Vries HJ, Geertzen JH, Reneman MF. Dose or content? Effectiveness of pain rehabilitation programs for patients with chronic low back pain: A systematic review. Pain. 2014 Jan;155(1):179-89.
24. Skouen JS, Grasdal AL, Haldorsen EM, Ursin H. Relative cost-effectiveness of extensive and light multidisciplinary treatment programs versus treatment as usual for patients with chronic low back pain on long-term sick leave: randomized controlled study. Spine (Phila Pa 1976). 2002 May 1;27(9):901-9; discussion 9-10.
25. van Dongen JM, van Wier MF, Tompa E, Bongers PM, van der Beek AJ, van Tulder MW, et al. Trial-based economic evaluations in occupational health: principles, methods, and recommendations. J Occup Environ Med. 2014 Jun;56(6):563-72.
26. Hlobil H, Uegaki K, Staal JB, de Bruyne MC, Smid T, van Mechelen W. Substantial sick-leave costs savings due to a graded activity intervention for workers with non-specific sub-acute low back pain. Eur Spine J. 2007 Jul;16(7):919-24.
27. Loisel P, Lemaire J, Poitras S, Durand MJ, Champagne F, Stock S, et al. Cost-benefit and cost-effectiveness analysis of a disability prevention model for back pain management: a six year follow up study. Occup Environ Med. 2002 Dec;59(12):807-15.
28. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA. 2012 Dec 26;308(24):2594-604.

29. Engel L. The Need for a New Medical Model: A Challenge for Biomedicine. Science. 1977;196(4286):129-36.

30. Head SJ, Kaul S, Bogers AJ, Kappetein AP. Non-inferiority study design: lessons to be learned from cardiovascular trials. Eur Heart J. 2012 Jun;33(11):1318-24.

31. Skouen JS, Grasdal A, Haldorsen EM. Return to work after comparing outpatient multidisciplinary treatment programs versus treatment in general practice for patients with chronic widespread pain. Eur J Pain. 2006 Feb;10(2):145-52.

32. Staal JB, Hlobil H, Twisk JW, Smid T, Koke AJ, van Mechelen W. Graded activity for low back pain in occupational health care: a randomized, controlled trial. Ann Intern Med. 2004 Jan 20;140(2):77-84.

33. Reneman MF, Beemster TT, Edelaar MJ, van Velzen JM, van Bennekom C, Escorpizo R. Towards an ICF- and IMMPACT-Based Pain Vocational Rehabilitation Core Set in the Netherlands. J Occup Rehabil. 2013 Feb 2.

34. Bouwmans C, Hakkaart-van Roijen L, Koopmanschap M, Krol M, Severens H, Brouwer W. Manual of the iMTA Productivity Cost Questionnaire (iPCQ). Rotterdam: 2013.

35. Hakkaart-van Roijen L, Tan S, Bouwmans C.Handleiding voor kostenonderzoek. Methoden en standaard kostprijzen voor economische evaluaties in de gezondheidszorg (Cost research guide. Methods and standard cost prices for economic evaluations in health care). Rotterdam: College voor zorgverzekeringen 2010.

36. Bouwmans C, De Jong K, Timman R, Zijlstra-Vlasveld M, Van der Feltz-Cornelis C, Tan SS, et al. Feasibility, reliability and validity of a questionnaire on healthcare consumption and productivity loss in patients with a psychiatric disorder (TiC-P). BMC Health Serv Res. 2013;13:217.

37. 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 Sep;36(5):404-12.

38. Soer R, Reneman MF, Vroomen PC, Stegeman P, Coppes MH. Responsiveness and minimal clinically important change of the Pain Disability Index in patients with chronic back pain. Spine (Phila Pa 1976). 2012 Apr 15;37(8):711-5.

39. Tait RC, Chibnall JT. Factor structure of the pain disability index in workers compensation claimants with low back injuries. Arch Phys Med Rehabil. 2005 Jun;86(6):1141-6.

40. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. Health Econ. 1993 Oct;2(3):217-27.

41. van der Zee K, Sanderman R. Het meten van de algemene gezondheidstoestand met de RAND-36. Een handleiding (Measuring the general health status with the RAND-36. A manual). Groningen: Noordelijk Centrum voor Gezondheidsvraagstukken, Rijksuniversiteit Groningen; 1993.

42. Lamers LM, McDonnell J, Stalmeier PF, Krabbe PF, Busschbach JJ. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ. 2006 Oct;15(10):1121-32.

43. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. Kwaliteit van leven meten in economische evaluaties: het Nederlands EQ-5D-tarief (Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff). Ned Tijdschr Geneeskd 2005, 149(28):1574-1578.

44. Hakkaart-van Roijen L, van Straten A, Donker M, Tiemens B. Manual Trimbos/iMTA questionnaire for Costs associated with Psychiatric illness (TiC-P). Rotterdam: 2002.

45. van den Brink M, van den Hout WB, Stiggelbout AM, Putter H, van de Velde CJ, Kievit J. Self-reports of health-care utilization: diary or questionnaire? Int J Technol Assess Health Care. 2005 Summer;21(3):298-304.

46. Selander J, Marnetoft SU, Bergroth A, Ekholm J. Return to work following vocational rehabilitation for neck, back and shoulder problems: risk factors reviewed. Disabil Rehabil. 2002 Sep 20;24(14):704-12.

47. van Duijvenbode DC, Hoozemans MJ, van Poppel MN, Proper KI. The relationship between overweight and obesity, and sick leave: a systematic review. Int J Obes (Lond). 2009 Aug;33(8):807-16.

48. Vendrig A. De vragenlijst arbeidsreïntegratie (The Work Reintegration Questionnaire). Diagnostiek-Wijzer 2005, 8:27-39.

49. Vendrig L, van Hove M, van Meijel M, Donceel P. Voorspellen van de verwachte verzuimduur met de Vragenlijst ArbeidsReïntegratie (VAR) (Predicting the expected duration of absence with the Work Reintegration Questionnaire (WRQ)). TBV 2011, 19(1):7-13.

50. Turner JA, Franklin G, Fulton-Kehoe D, Sheppard L, Wickizer TM, Wu R, et al. Worker recovery expectations and fear-avoidance predict work disability in a population-based workers' compensation back pain sample. Spine (Phila Pa 1976). 2006 Mar 15;31(6):682-9.

51. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. Pain. 1986 Oct;27(1):117-26.

52. Soer R, van der Schans CP, Geertzen JH, Groothoff JW, Brouwer S, Dijkstra PU, et al. Normative values for a functional capacity evaluation. Arch Phys Med Rehabil. 2009 Oct;90(10):1785-94.

53. Kuijer PP, Gouttebarge V, Brouwer S, Reneman MF, Frings-Dresen MH. Are performance-based measures predictive of work participation in patients with musculoskeletal disorders? A systematic review. Int Arch Occup Environ Health. 2012 Feb;85(2):109-23.

54. Schumi J, Wittes JT. Through the looking glass: understanding non-inferiority. Trials. 2011;12:106.

55. Krol M, Brouwer W. How to estimate productivity costs in economic evaluations. Pharmacoeconomics. 2014 Apr;32(4):335-44.

56. Koopmanschap MA, Rutten FF, van Ineveld BM, van Roijen L. The friction cost method for measuring indirect costs of disease. J Health Econ. 1995 Jun;14(2):171-89.

57. Campbell MK, Torgerson DJ. Bootstrapping: estimating confidence intervals for cost-effectiveness ratios. QJM. 1999 Mar;92(3):177-82.

58. Fenwick E, O'Brien BJ, Briggs A. Cost-effectiveness acceptability curves--facts, fallacies and frequently asked questions. Health Econ. 2004 May;13(5):405-15.

59. Anema JR, Jettinghoff K, Houtman I, Schoemaker CG, Buijs PC, van den Berg R. Medical care of employees long-term sick listed due to mental health problems: a cohort study to describe and compare the care of the occupational physician and the general practitioner. J Occup Rehabil. 2006 Mar;16(1):41-52.

60. Steenstra IA, Verbeek JH, Prinsze FJ, Knol DL. Changes in the incidence of occupational disability as a result of back and neck pain in the Netherlands. BMC Public Health. 2006;6:190.

61. van Oostrom SH, Anema JR, Terluin B, de Vet HC, Knol DL, van Mechelen W. Cost-effectiveness of a workplace intervention for sick-listed employees with common mental disorders: design of a randomized controlled trial. BMC Public Health. 2008;8:12.

62. Vermeulen SJ, Anema JR, Schellart AJ, van Mechelen W, van der Beek AJ. Cost-effectiveness of a participatory return-to-work intervention for temporary agency workers and unemployed workers sick-listed due to musculoskeletal disorders: design of a randomised controlled trial. BMC Musculoskelet Disord. 2010;11:60.

63. Gorringe J. Initial preparation for clinical trials. In: Harris E, Fitzgerald J, editors. The principles and practice of clinical trials. Edinburgh: Livingstone; 1970.

64. Wet verbetering poortwachter (The Law on Gate keeping Disability Insurance). Accessed 22 January 2018; Available from: https://www.arboportaal.nl/onderwerpen/wet-verbetering-poortwachter.

65. Committee for Medicinal Products for Human Use (CHMP) guideline on the choice of the non-inferiority margin. Statics in Medicine. 2006;25(10):1628-38.

66. Schaafsma F, Schonstein E, Whelan KM, Ulvestad E, Kenny DT, Verbeek JH. Physical conditioning programs for improving work outcomes in workers with back pain. Cochrane Database Syst Rev. 2010(1):CD001822.

67. Severens JL, Mulder J, Laheij RJ, Verbeek AL. Precision and accuracy in measuring absence from work as a basis for calculating productivity costs in The Netherlands. Soc Sci Med. 2000 Jul;51(2):243-9.

68. Karrholm J, Ekholm K, Ekholm J, Bergroth A, Ekholm KS. Systematic co-operation between employer, occupational health service and social insurance office: a 6-year follow-up of vocational rehabilitation for people on sick-leave, including economic benefits. J Rehabil Med. 2008 Aug;40(8):628-36.

69. Lotters F, Meerding WJ, Burdorf A. Reduced productivity after sickness absence due to musculoskeletal disorders and its relation to health outcomes. Scand J Work Environ Health. 2005 Oct;31(5):367-74.

70. Koopmanschap M, Burdorf A, Jacob K, Meerding WJ, Brouwer W, Severens H. Measuring productivity changes in economic evaluation: setting the research agenda. Pharmacoeconomics. 2005;23(1):47-54.

71. Hamer H, Gandhi R, Wong S, Mahomed NN. Predicting return to work following treatment of chronic pain disorder. Occup Med (Lond). 2013 Jun;63(4):253-9.

72. Bannink F. Oplossingsgerichte vragen. Handboek oplossingsgerichte gespreksvoering (Solution-oriented questions. Solution-oriented interview manual), 2nd edn: Pearson Benelux B.V. 2009.

73. Green L, Oades L, Grant A. Cognitive-behavioral, solution-focused life coaching: Enhancing goal striving, well-being, and hope. The Journal of Positive Psychology. 2006;1(3):142-9.

74. De Koning J. Empowerment and the empowering environment: EQUAL, European Commission 2007.

75. Lindstrom I, Ohlund C, Eek C, Wallin L, Peterson LE, Fordyce WE, et al. The effect of graded activity on patients with subacute low back pain: a randomized prospective clinical study with an operant-conditioning behavioral approach. Phys Ther. 1992 Apr;72(4):279-90; discussion 91-3.

76. Lindstrom I, Ohlund C, Nachemson A. Physical performance, pain, pain behavior and subjective disability in patients with subacute low back pain. Scand J Rehabil Med. 1995 Sep;27(3):153-60.

77. Steenstra IA, Anema JR, Bongers PM, de Vet HC, van Mechelen W. Cost effectiveness of a multi-stage return to work program for workers on sick leave due to low back pain, design of a population based controlled trial [ISRCTN60233560]. BMC Musculoskelet Disord. 2003 Nov 21;4:26.

Appendix 1. Intervention modules

Work participation coordination Work participation coordination is carried out by a work participation coordinator [71], and encompasses a workplace visit, case management, and two evaluation moments. The workplace visit includes an ergonomic workplace analysis and a consultation with employer and employee (patient), with the aim of developing a work participation plan. Case management consists of individual coaching sessions with the patient and unplanned ad hoc conversations with the patient during the programme. The coaching style of the work participation coordinator is mainly based on solution-focused coaching [72, 73] and on empowerment [74]. The two evaluation moments include a report on progression in work participation, which will be performed at mid-term and discharge. Dosage: 10 hours in both programmes.
Graded activity Graded activity is based on the protocol designed by Lindstrom [75, 76], and adjusted to the Dutch situation [8, 32, 77]. The graded activity programme is carried out by a physical therapist. The purpose of graded activity is to restore occupational functioning and to facilitate work participation. During the programme, the patient has an active role and the physical therapist acts as a coach and supervisor, using a hands-off approach [77]. Graded activity is a time-contingent approach with an increase in load and complexity of movements. To attain physical reconditioning, the graded activity protocol may be supplemented with endurance exercises. Dosage: 60 hours (2 x 2 hours per week) for patients in the 100-h VR group. The amount of graded activity in the 40-h VR group will differ per patient.
Cognitive behavioural therapy Cognitive behavioural therapy (CBT) is carried out by a psychologist and consists of individual sessions, group education, and unplanned ad hoc conversations during the programme. The CBT sessions are based on solution-focused coaching [72, 73] and empowerment [74]. It encompasses items such as coping, cognition, communication, and self control. Dosage: the dosage differs per patient, but we factor in approximately 30 minutes per week for patients in the 100-h VR group, resulting in a total of 7.5 hours. The amount of CBT in the 40-h VR group will differ per patient.
Group education Group education encompasses physical and mental topics, and will be carried out by a physical therapist and a psychologist. Physical topics are the effect of physical activity on the body (i.e. training principles), chronic and acute pain, pain sensitization, anatomy and ergonomics, and nutritional recommendations pre- and post exercise. Mental topics are empowerment, setting graded tasks, cognitive behavioural therapy, and coping with pain. Dosage: 15 hour (60 minutes per week) for patients in the 100-h VR group. The amount of group education in the 40-h VR group will differ per patient.
Relaxation Relaxation sessions are carried out by a physical therapist. Different techniques are employed, such as meditation, visualization, autogenic training, mindfulness, breath control, progressive relaxation, and reciprocal inhibition. The aim of relaxation is improved body awareness and experiencing the difference between tension and relaxation of the muscles. Dosage: 7.5 hours (30 minutes per week) for patients in the 100-h VR group. The amount of relaxation in the 40-h VR group will differ per patient.

CHAPTER 6

Usefulness and Feasibility of Comprehensive and Less Comprehensive Vocational Rehabilitation for Patients with Chronic Musculoskeletal Pain: Perspectives from Patients, Professionals, and Managers

T. Beemster
J. van Velzen
C. van Bennekom
M. Reneman
M. Frings-Dresen

Submitted

Abstract

Purpose

The purpose of this study was to explore the usefulness and feasibility of comprehensive vocational rehabilitation (C-VR) and less comprehensive VR (LC-VR) for workers on sick leave due to CMP.

Materials and methods

Semi-structured interviews were held with patients, professionals, and managers. Using topic lists, participants were questioned about barriers to and facilitators of the usefulness and feasibility of C-VR and LC-VR. All interviews were transcribed verbatim. Data were analyzed by systematic text condensation using thematic analysis.

Results

Thirty interviews were conducted with thirteen patients (n=6 C-VR, n=7 LC-VR), eight professionals, and nine managers. Three themes emerged for usefulness (“patient factors”, “content”, ”dosage”) and six themes emerged for feasibility (”satisfaction”, ”intention to continue use”, ”perceived appropriateness”, ”positive/negative effects on target participants”, ”factors affecting implementation ease or difficulty”, ”adaptations”).

Conclusions

The patients reported that both programs were feasible and generally useful. The professionals preferred working with the C-VR, although they disliked the fixed and uniform character of the program. They also mentioned that this program is too extensive for some patients, and that the latter would probably benefit from the LC-VR program. Despite their positive intentions, the managers stated that due to the Dutch healthcare system, implementation of the LC-VR program would be financially unfeasible.

Keywords

Qualitative research, Tailored intervention, Workplace intervention, Return to work, Implementation research.

Introduction

Chronic musculoskeletal pain (CMP) is a common condition that contribute to disability, a decline in work participation, and substantial costs [1, 2]. Multidisciplinary bio-psychosocial programs, such as vocational rehabilitation (VR), are advocated to enhance the work participation of sick-listed workers with CMP [3, 4]. VR is defined as "a multiprofessional evidence-based approach to optimize work participation that includes various services and activities provided in different settings to working age individuals with health related impairments, limitations, or restrictions in work functioning" [5]. A review found that working-age adults on sick leave with musculoskeletal disorders who received VR saved 40 days of sick leave at twelve months follow-up compared to care as usual [6]. Another review showed that VR saved 1.11 (interquartile range 0.32-3.20) sick-leave days per month compared to care as usual [7].

In general, the content of VR programs covers three bio-psychosocial domains: a) health-focused (i.e., health services intervention subcategories, such as graded activity/exercise, cognitive behavioral therapy (CBT), education, work-hardening); b) service coordination (i.e., improving communication within the workplace or between the workplace and healthcare providers); and c) work modification (i.e., modified duties, modified working hours, supernumerary replacements, ergonomic adjustments or other worksite adjustments) [8]. Some modules are executed in a group, such as education and CBT, and others are executed in a one-to-one setting, such as sessions with a case manager or psychologist. Nonetheless, VR programs can vary widely in terms of content [4, 9], and it is unclear how many contact hours of each type of content are necessary to achieve the best results [7, 9-12]. The latter issue is illustrated by a review that showed that effective multidisciplinary VR programs for patients with CMP ranged from fewer than six contact hours to more than 70 contact hours [7]. Another review showed that pain rehabilitation programs ranging from seven to 197 contact hours were effective in enhancing the work participation of patients with CMP [9]. Furthermore, three randomized controlled trials (RCTs) showed that VR programs with different numbers of contact hours (18.5-h vs 52-h [13], 15-h vs 120-h [14], 10-h vs 120-h) [15, 16], respectively, were non-inferior to each other with regard to enhancing the work participation of sick-listed workers with CMP.

Remarkably, despite growing evidence that less comprehensive VR (LC-VR) might be non-inferior compared to comprehensive (C-VR), little uptake has

been observed in clinical practice, apart from controlled studies. One possible explanation for this is that VR is often complex [18], as it consists of many elements, involves many stakeholders, and is embedded in an administrative, financial, and social context [17]. When implementing a new intervention in clinical practice, it is recommended that the opinions of patients, professionals, managers, and policymakers regarding the feasibility and usefulness of the intervention are taken into account [17, 19, 20]. Usefulness is defined as the suitability of an intervention for the intended purpose and the extent to which it meets the needs of important users [21]. It can encompass three dimensions: usefulness on an individual level, on an organizational level, and of the intervention itself [21]. Feasibility studies can help us to evaluate and prioritize whether or not it will be feasible to conduct a new intervention, and whether all the necessary components of the new intervention will work together effectively [19, 22]. The feasibility of an intervention can encompass different areas, such as the satisfaction of target participants, the appropriateness of the intervention for patients, the effect of the intervention on the organization, the effect of the intervention on participants, implementation factors, and adaptations [19].

In the Netherlands, a number of rehabilitation centers perform care-as-usual multidisciplinary C-VR programs of ~100 contact hours. The C-VR program consists of health-focused modules (fitness/graded activity, CBT, group education, and relaxation) and return to work (RTW) coordination (service coordination and work modifications). In an RCT, the C-VR program was compared with a less comprehensive program (LC-VR) of ~40 contact hours [23]. The LC-VR program comprised a fixed part (RTW coordination) and a tailored part consisting of individually-chosen components of the C-VR program's health-focused modules. The RCT was conducted between November 2014 and January 2016 (more information about the RCT is provided in a study protocol paper [23]). As the necessary inclusion rate was hampered, however, the study was discontinued. Nonetheless, eight patients completed the LC-VR program and six patients completed the C-VR program. The aim of this paper is to explore the usefulness and feasibility of a C-VR program and a LC-VR program for workers on sick leave due to chronic musculoskeletal pain, from the perspective of patients, professionals, and managers.

Materials and methods

The consolidated criteria for reporting qualitative research (COREQ) checklist was used when designing the study [24].

Participants

For this qualitative study, three groups of stakeholders were interviewed: i) patients who had completed the LC-VR or C-VR; ii) professionals who had executed at least one LC-VR program and who had several years of experience with the C-VR program; and iii) managers from centers who had executed the LC-VR and the C-VR programs, and managers from centers who had executed the C-VR program alone. The latter were included in this study in order to enrich our understanding of program feasibility.

The vocational rehabilitation programs

Comprehensive vocational rehabilitation
The comprehensive vocational rehabilitation (C-VR) program was a multidisciplinary bio-psychosocial group-based program that consisted of five modules: RTW coordination, fitness/graded activity, CBT, group education, and relaxation. RTW coordination consisted of service coordination (communication part: individual sessions with the patient, conduct a RTW plan, and a workplace visit, including a conversation with the patient and supervisor/employer) and work modifications (ergonomic part). A detailed description of the content of the C-VR program can be found elsewhere [23]. The C-VR program covered approximately 100 contact hours and lasted fifteen weeks, with two contact moments of approximately 3.5 h/session each week.

Less comprehensive vocational rehabilitation
The less comprehensive vocational rehabilitation (LC-VR) program was a multidisciplinary bio-psychosocial group-based program that consisted of a fixed part (RTW coordination, ~10 hours) and a tailored part (~30 hours). The content of the tailored part was based on a VR-team decision taken after a multidisciplinary screening; only those modules that were deemed most useful were chosen. The LC-VR program covered a maximum of 40 hours over fifteen weeks. In general, the program was based on the following blueprint: weeks 1-5, two sessions/week; weeks 6-10, one session/week; weeks 11-15, 2-3 sessions in five weeks. Professionals were free to change this blueprint.

Context

The stakeholders in this study fell under the Dutch sickness compensation and healthcare system. When an employee is sick-listed in the Netherlands, both the employee and employer are responsible for the work participation process during the first two years of sick leave. According to the Dutch Gatekeeper Improvement Act, the employer has to provide wage replacement during this two-year period [25]. If VR is indicated for the employee and a workplace intervention is needed, the cost of this module (approximately €1,200) must be reimbursed by the employer. Other aspects of VR (i.e., fitness, CBT, relaxation therapy, group education, etc.) are reimbursed by healthcare insurers. The amount that is reimbursed is categorized stepwise and depends on a number of reimbursement factors, such as program duration, group size, the number of professionals in a group, whether it is individual or group care, and so forth. In particular cases, several additional hours or weeks can make a difference in program reimbursement of thousands of euros.

Data collection

Semi-structured telephone interviews were conducted between the interviewer and stakeholders; non-participants were absent. The interviews were held between June and October 2016. All interviews were held by TB (male, exercise therapist, health scientist, PhD candidate, participated in a course on conducting Qualitative Health Research). Thirty-two interviews were planned: patients n=14 (LC-VR: n=8; C-VR n=6), professionals n=8 (two per center), and managers n=10 (experiences with LC-VR and C-VR: n=4; experiences with C-VR: n=6). Topic lists were used as a framework for the interviews; these lists included topics on the usefulness and feasibility of the LC-VR and C-VR programs. Logical reasoning was used to develop the usefulness topics, while the feasibility topics were derived from a range of sources [19, 26, 27]. The patients and professionals were questioned about the usefulness, feasibility, barriers to and facilitators of both programs. The managers were asked about feasibility, barriers to and facilitators of the program(s). The professionals and managers were asked about a hypothetical situation in which the LC-VR program was implemented as the new care-as-usual program and the C-VR program was continued as the care-as-usual program. Patients were asked to indicate their satisfaction with the allocated program on a 0-10 scale (0=not satisfied at all, 10=very satisfied). Patients were also asked to evaluate the usefulness of each program module. Two pilot interviews were performed (with a professional and a manager) to test

the topic list and to train the interviewer in the interview process. After this pilot phase, the final topic lists were produced (Appendix 1). A few days before each interview, an e-mail and a letter with information about the interview were sent to each stakeholder. The letter explained that the interview was confidential, and asked for permission to audiotape the interview and save the audio file and transcription for fifteen years. This storage time is in accordance with the institutional research code [28]. Before each interview, the same information was repeated and informed consent was given. The patients had already given their written informed consent as part of an RCT [23] and the professionals and managers gave their consent verbally before the start of the interview. Participants were asked to state their opinions openly, and it was explained that there were no good or bad answers. After completion of the interviews, field notes were written down as soon as possible. The field notes consisted of descriptive information such as the date and time, setting, action, behavior, and conversations observed; and reflective information such as thoughts, ideas, questions, and concerns raised in the interview. Patients' characteristics were obtained from baseline questionnaires from an RCT [23].

Data analyses

All interviews were audiotaped and transcribed verbatim. The interviews were transcribed by the interviewer and an assistant, and all transcriptions were verified and corrected by the first author. Data were analyzed by systematic text condensation using theoretical thematic analysis, a method for identifying, analyzing, and reporting themes within data [29]. The analysis was performed in a series of five steps: (1) familiarization with the data; (2) generation of initial codes; (3) searching for themes; (4) defining and naming themes; and (5) producing the report [29]. Three transcriptions per stakeholder were analyzed in duplicate (patients: first author and fourth author MR; professionals: first author and second author JVV; managers: first author and last author MFD). The codes and themes that emerged from the data were compared and discussed until consensus on a preliminary set of labels was reached. The final interviews were analyzed by the first author TB. Consensus was reached with all authors about a final code tree (a set of themes and codes). The report was produced with reference to the areas of feasibility used by Bowen et al. [19]. The interviews were analyzed using the computer software program MAXQDA version 12 (VERBI Software. GmbH Berlin, Germany 2015).

Results

Participants

In the present study, nine managers, eight professionals, and thirteen patients were interviewed. The response rate of the interviews was 30 out of 32 participants. One manager refused to participate because he was working on an interim basis, and one patient refused to participate. Of the patients, n=7 out of 8 had participated in the LC-VR program and n=6 out of 6 had participated in the C-VR program. The general participant characteristics are shown in table 1.

Table 1. Baseline characteristics (age, gender, education, sick-leave status, pain) of patients participating in this study

	C-VR	N	LC-VR	N
Age, years (mean, SD)	47 (11)	6	44 (14)	7
Gender (% female)	100	6	43	3
Education (% low)	17	1	29	2
Sick-leave status (%) [a]				
No sick leave (working fulltime)	0	0	14	1
Part-time sick leave	50	3	29	2
100% sick leave	50	3	57	4
Pain duration (%)				
<6 months	17	1	57	4
>6 months	83	5	43	3
Pain mean 0-10 (mean, SD) [b]	6.0 (0.6)	6	6.4 (1.9)	7
Pain worse 0-10 (mean, SD) [c]	7.8 (0.8)	6	8.3 (1.3)	7

C-VR, comprehensive vocational rehabilitation; LC-VR, less comprehensive vocational rehabilitation; SD, standard deviation.
[a] Obtained with the question: 'Are you working full-time at the moment?' Answer categories: 'Yes', 'No, I am working part-time', 'No, I am on full-time sick leave'.
[b] Pain on average in the preceding week: 0=no pain, 10=worst possible.
[c] When the pain was worst in the preceding week: 0=no pain, 10=worst possible pain.

Interviews

In total, 30 participants were interviewed and analyzed. The interviews lasted 16-46 minutes (mean 27 ± 7 minutes), excluding the introduction time. When the final interviews were analyzed, we saw the same categories, rather than new categories, indicating data saturation.

Themes

Code trees for usefulness and feasibility were developed (Appendices 2-3). From these, three themes emerged for usefulness (”patient factors”, ”content” and ”dosage”), and six themes emerged for feasibility (“satisfaction”, ”intention to continue use”, ”perceived appropriateness”, ”positive/negative effects on target participants”, ”factors affecting implementation ease or difficulty”, and ”adaptations”). When describing the results, codes were placed in bold, and statements by the three actors were abbreviated as PT (patients), PR (professionals), and MA (managers).

Usefulness

Patient factors
The professionals mentioned that the LC-VR program was useful for some of the patients referred to VR, but not for all of them.

✓ I think that it’s suitable for some and not for others (21, PR).

However, they also stated that the C-VR program did not suit all patients, either.

✓ I expect that it [LC-VR] would indeed be good for a certain group, but there are also people who, well, who need slightly more intensive guidance [C-VR] (26, PR).

To guide which program would be useful for which ”type” of patients, the professionals mentioned various patient factors. These were clustered into five categories (intelligence, behavioral, complaints, mental, and work) and 25 codes (Codes and Quotations: Table 2).

Table 2. Patient-related factors determining the usefulness of C-VR or LC-VR according to professionals

Category	Patient-related factors determining the usefulness of the C-VR program	Patient-related factors determining the usefulness of the LC-VR program	ID	Quotation
Intelligence	Low level of education	High level of education	28, PR	LC-VR: The people who are already more proactive, who are slightly more independent, perhaps further on in the process, too, and a bit more highly educated.
	Low level of knowledge	High level of knowledge	28, PR	LC-VR: People who have a lot of knowledge and insight, who can process things more quickly and who are able to change themselves a bit.
	Low level of information uptake	High level of information uptake	26, PR	LC-VR: People, I think, who are also able, yes, to pick things up quickly, who are perhaps more independent in that sense.
Behavioral	Not proactive person	Proactive person	29, PR	LC-VR: I think, especially people with a very proactive coping style, who are, um, rapidly encouraged to take charge of things.
	Low level of self-direction	High level of self-direction	21, PR	LC-VR: People who simply find it difficult to take charge of the process, the rehabilitation process, for them, I think 40 hours is too little.
	Low level of discipline	High level of discipline	22, PR	LC-VR:... it requires a degree of discipline to pick things up at home or in any case from home, such as sports, and to also apply other things, that takes discipline.
	Low level of willingness to change [a]	High level of willingness to change	30, PR	LC-VR: Willingness to change, looking at themselves, that kind of factors.
	Patient can not train independently	Patient can train independently	21, PR	LC-VR: We also see a lot of people who are not really able to work out independently.
Complaints	Fibromyalgia	No fibromyalgia [a]	29, PR	C-VR: I think for example fibromyalgia or chronic fatigue like ⋯
	Chronic fatigue	No chronic fatigue [a]	29, PR	C-VR: I think for example fibromyalgia or chronic fatigue like ⋯
	Chronic complaints [a]	Subacute complaints	24, PR	LC-VR: Someone who hasn't been out for very long ... who's at a very early stage in the process ... yes, the subacute or when C-VR is used as a prevention program.
	Low capacity [b]	High capacity	28, PR	C-VR: Those who, when it comes to taking things on, mentally and physically, have so little resilience that they first have to build up a certain degree of strength before they are able to do anything at all meaningful at work.

Table 2. (Continued)

Category	Patient-related factors determining the usefulness of the C-VR program	Patient-related factors determining the usefulness of the LC-VR program	ID	Quotation
	High / much psychosocial problems	No / low psychosocial problems	25, PR	LC-VR: ··· where less psychosocial problems play a role.
			25, PR	C-VR:... then you see that it really is a very considerable problem and yeah, that it's therefore not only a work problem, but also a psychosocial problem, one that's often very, very complicated, too.
	Multi-problems	No multi-problems [a]	21, PR	C-VR: ... where, for example, there are problems on multiple levels, where there are problems at work and at home, yes, how to put it, intrinsically, so it's more that people are coming up against their own difficulties, yes, good. On a psychological level, but also in dealing with and accepting their symptoms.
	Complex patients	No complex patients [a]	24, PR	C-VR:... that people have often tried something else and when that really hasn't worked, then the [name of center] comes up, so yes, if it doesn't work there anymore, then... you know.
Mental	Movement anxiety	No movement anxiety	28, PR	C-VR: People with a lot of anxiety associated with movement, who just need a little more attention to be able to overcome that anxiety too.
			29, PR	LC-VR: If they nevertheless dare to train at the gym, while they are afraid. Still dare to train, even if it's painful, then some people will conclude more quickly, um, OK, I can do it, so I'll start working on my development.
	Obstructive thoughts	No obstructive thoughts [a]	29, PR	C-VR: It depends on the extent of that obstructive thought. As a psychologist, yeah, that's something you can't express in numbers, say, but you can talk about certain gradations. Um, let's think, for example, I now have someone training and, um, now, well, that one frets a little about pain and fatigue, but other people really fret day in, day out, and then it obstructs them much more in their daily life. So there's a difference of gradation there. And the degree of gradation also determines how much work you have to put in in order, um, to reduce that gradation.
	Uncertain patients	Confident patients [a]	28, PR	C-VR: I think people who generally chose the C-VR program, that people are what I just said, who feel pretty insecure.

Table 2. (Continued)

Category	Patient-related factors determining the usefulness of the C-VR program	Patient-related factors determining the usefulness of the LC-VR program	ID	Quotation
	Low cognitions	High cognitions [a]	25, PR	LC-VR: ... yet the problem around his pain experience and his cognitions were much stronger, and that we did not get that turned around in the LC-VR program, not even a start with that.
	Low acceptance of complaints	Acceptance of complaints [a]	21, PR	LC-VR: ⋯ but also in dealing with and accepting their complaints.
Work factors	Work participation not treatment goal [a]	Work participation as the treatment goal	25, PR	LC-VR: ... where it really concerns a work-related question.
	No willingness to return to work [a]	Willingness to return to work	25, PR	LC-VR: ... someone who's also more open to it... like, I want to do this quickly and I also want to get back to work quickly.
	Has not made steps towards work reintegration [a]	Has made steps towards work reintegration	25, PR	LC-VR: ⋯ who has already taken some steps in the direction of work.
	Bad relationship with employer	Good relationship with employer [a]	29, PR	C-VR: people who have a worse relationship with the employer.
	Long time off work	≤ one year off work [a]	30, PR	C-VR: people with more long-term symptoms, that is, people who may have been on sick leave for over a year.
			25, PR	C-VR: sometimes they've been at home for even longer, meaning they've been out of the work environment for longer, perhaps then it all gets worse in their head, so they're no longer able to pick up the thread, yes, I think that could really be one of the factors.

C-VR, comprehensive vocational rehabilitation; LC-VR, less comprehensive vocational rehabilitation; NM, not mentioned.
[a] Patient-related factor which was not explicitly mentioned by professionals but rather indirect (implicit). For example: the codes 'Low level of willingness to change' and 'High level of willingness to change'[3], were mentioned as useful (i.e., high level) and not useful (i.e., low level) patient factors for the LC-VR program, but were not explicitly mentioned as a patient-related factor determining the usefulness of the C-VR program. The professionals however implicitly mentioned that such not useful patient-related factors for the LC-VR program (in this example low level of willingness to change) was in fact an eligible (useful) patient factor for the C-VR program.
[b] General capacity, mental capacity, and physical capacity together

[3] See also Appendix 2.

Content

Some patients reported that they had found all of the content useful, i.e., the **whole program**; and some patients mentioned that some content had been partly useful and/or not useful (Table 3). In addition, some patients stated that the group education sessions and sessions with psychologist had been useful at the start of the program, but not at the end (i.e., content saturation):

✓ I found it useful, but at a certain point, it all became much of a muchness, if you know what I mean. At a certain point, you know what kind of pain Peter has and what kind of pain Paul has (694, LC-VR).

In contrast, some patients said that the relaxation sessions had not been useful at the start of the program, but they had been useful at the end:

✓ Um ... eventually, yes. In the beginning, I thought it was really bad. I felt like, 'What am I doing here?' (489, C-VR).

Table 3. Usefulness of the content of the C-VR and LC-VR programs, as mentioned by patients

	C-VR		LC-VR	
Content	Useful content	Not useful content [a]	Useful content	Not useful content [a]
Relaxation	x	x	x	x
Fitness	x		x	
Psychologist	x	x	x	x
Group education	x	x	x	x
RTW coordination - ergonomic part	x	x	x	x
RTW coordination - communication part	x		x	
Movement teacher [b]	x		x	
Aquatic exercises [c]	x		x	

C-VR, comprehensive vocational rehabilitation; LC-VR, less comprehensive vocational rehabilitation; RTW, return to work
[a] 'Partly useful' and 'Not useful' taken together
[b] Undertaken at two centers
[c] Undertaken at one center

Dosage
The patients stated that the dosage of the program they had followed was feasible.

- ✓ Yes, it was easy. Yes, I had to go along two mornings a week, and yes, in principle I also got time off work (313, LC-VR).

- ✓ I went twice a week, yeah, so my employer gave me the chance to go along (605, C-VR).

Among the patients, however, there was a wide range of opinions about the optimal **dosage** of the program (if they had the chance to change it). The statements about program dosage were similar for both programs (Appendix 4).

Concerning the usefulness of the dosage of the C-VR program, the patients and the professionals agreed that no more treatment hours were needed to achieve better results. In fact, it was suggested that **the C-VR program could be slightly shorter** (PT, PR), and that **less complex patients would probably benefit from a shorter program** (PR, MA). On the other hand, some professionals stated that having 100 hours gave them enough space to deliver **tailored care**, and **enough time** to perform **physical training principles**, achieve **behavioral change**, **explain the sensitization story**, encourage patients to **take up healthy behavior**, **explore extra interventions,** deal with the appearance of **an unforeseen co-morbidity**, or **build a relationship with the employer** (Quotations: Appendix 4).

Concerning the usefulness of the dosage of the LC-VR program, there was a discrepancy between patients and professionals. On the one hand, the patients stated that the dosage they received was appropriate to achieve their treatment goal(s). On the other hand, the professionals stated that the dosage of the LC-VR program was generally **too low** for the majority of people who are referred to VR.

- ✓ I think that 40 hours is very tight if you really want to change behavior. I wonder whether it's feasible, now I've done it like that twice and also kept more of an eye on how it's done. I think it's very tight (24, PR).

Feasibility

Satisfaction

Patients **rated** the LC-VR program as positive (mean: 8, min-max: 7-9), and the C-VR program as positive (mean: 7.8, min-max: 4.5-9). Patients had **positive** and **negative experiences** with both programs:

- ✓ I'm really satisfied, yeah. I'm extremely satisfied, it did me a lot of good (476, C-VR).

- ✓ I found that from the beginning, quite a bit was said about the fact that, yeah, it might all be in your head, if you've been in pain for that long you think you're still in pain, and in my case, I didn't believe that beforehand, and hearing that there might be no treatment for you left or there not being any other options, yeah, I simply didn't know about that, so when I began, I thought that I really would get better and would also be able to do more, and during the course I found that, if I said I really was in pain and that I wasn't able to do things properly, that it was often ignored (696, C-VR). *Note: this patient left the program early because a serious medical problem appeared that had not previously been detected*.

- ✓ I'm certainly satisfied, I got lots out of it and learned loads (313, LC-VR).

- ✓ No, because I think I did it, of course, in the hope that it would get better, but OK, it didn't work out, even though I did all the exercises. I did it at home, too, I was also given little exercises to do, I did them all properly. (...) one explanation is that I probably have arthrosis all over my body, wear, I have it everywhere (212, LC-VR). *Note: this patient switched to the C-VR program because he/she had not achieved his/her treatment goals. However, patient did not achieve his/her treatment goals in the C-VR program, either*.

Professionals had **positive** and **negative experiences** with the LC-VR program:

- ✓ I think it's useful in that sense, because you look very specifically at, well, what's important for this client, so you really, so you make the patient dependent, and that, in any case, someone doesn't get something that they don't need so much, and what I also found kind of useful was that

the client takes charge of doing things at an earlier stage, which means that we're spoon-feeding them a bit less (26, PR).

✓ We took him/her out of the trial at a certain point, because we saw that, and coincidentally, another specific diagnosis was also made, so he/she had to go, but we were also very pleased that he/she went, because we actually needed more time (24, PR).

All of the professionals were **positive about the C-VR program.**

Intention to continue use
The patients stated that they would follow the program (LC-VR, C-VR) again if it proved necessary, and that they would **recommend** the program to family, friends, colleagues, etcetera, if necessary. The professionals preferred to continue using the C-VR program in clinical practice. Some professionals and managers would be **willing to work with the LC-VR program in the future**, if there were resources for this and adaptations were made (see "Factors affecting ease or difficulty of implementation" and "Adaptations"). One manager (from a non-participating RCT center) would be willing to implement the LC-VR program (or a similar program) as his/her new care-as-usual program. Another manager, also from a non-participating RCT center, would be willing to continue using the LC-VR program, since his/her center recently implemented a similar program (Quotations: Appendix 5).

Perceived appropriateness
The professionals mentioned that one single program (i.e., LC-VR or C-VR) would not be useful, and thus not appropriate, for all patients referred to VR. However, the professionals described the **C-VR program as the most appropriate program** for patients referred to VR, for the following reasons: having **enough time** (Appendix 4), because the C-VR program was the **current** and thus "**known**" **program** (for both professionals and referrers), for **logistical reasons**, and because the program is **financially beneficial** (Quotations: Appendix 5).

Positive and negative effects on target participants
Positive aspects of the LC-VR program were associated with the **dosage** (time schedule) of the program, such as spending **less time absent from work** (PR, MA), the prevention of **therapy dependency** (PR, MA), and increasing patient **self-management** (PT, PR). All actors mentioned **tailored care** as a positive

aspect of the LC-VR program. A final positive factor was **rehabilitation in a group** (PT).

Tailored care (PT, PR) and **rehabilitation in a group** (PT, PR) were also mentioned as positive effects of the C-VR program. The negative effects of the C-VR program included the creation of **therapy dependency** (PR, MA), and the fact that one is not forced to **think critically** about which content a patient really needs (PR). A further negative effect was **redundant care** (i.e., partly/not useful content) (PT, PR, MA) and as a consequence of this, the fact that the program is too **uniform** (PR, MA) (Quotations: Table 4).

Factors affecting ease or difficulty of implementation

Proper **reimbursement** of the LC-VR program was mentioned as being of paramount importance (PT, MA). The reimbursement of the **RTW coordination** module was stated as a key implementation factor (PR, MA), as well as avoiding **too much diversity** in the LC-VR program (PR). Another implementation factor was that the two programs should be delivered **separately** (PR, MA). The negative implementation factors for the LC-VR program included a **lack of evidence** (PR, MA) and **best practices** (MA), and the **prejudice** of professionals. The **rigid financial structure** of the Dutch healthcare system (which is unclear and can differ from year to year) was frequently mentioned as a negative factor for both programs (PR, MA) (Quotations: Appendix 5).

Adaptations

Patients, professionals, and managers suggested several adaptations with regard to content and delivery that they thought would optimize the LC-VR and/or C-VR program (Codes and Quotations: Appendix 6).

Discussion

This study provided insights into the usefulness and feasibility of C-VR and LC-VR for patients with CMP and reduced work participation, from the perspective of patients, professionals, and managers.

Usefulness

Five categories of patient factors (intelligence, behavioral, complaints, mental and work) were identified from the interviews with professionals. As suggested above, these patient factors could indicate which program would be useful for which type of patient. Our findings on the ”behavioral”, ”complaints”, and ”mental” patient factors were consistent with the findings of other qualitative studies assessing patients’ case complexity [30-32]. ”Intelligence” (i.e., high level of education) [33] and ”work” [4, 33, 34] were predicting factors for RTW in other studies.

A further “usefulness theme” in the present study concerned the content of the programs: a homogeneous pattern of’ ”useful”, ”partly useful”, and ”not useful” content emerged for the two programs. The findings on content are in line with those of other studies [4, 7, 8], showing that bio-psychosocial multidisciplinary (VR) programs are effective for people with CMP and impaired work participation. More specifically, a review has shown that implementing a multi-domain intervention with components in at least two of the following three domains – health-focused (i.e., health services intervention subcategories such as graded activity/exercise, CBT, work-hardening), service coordination (i.e., improving communication within the workplace or between the workplace and the healthcare providers), or work modification (i.e., modified duties, modified working hours, supernumerary replacements, ergonomic adjustments or other worksite adjustments) – can help reduce time lost as a result of musculoskeletal and pain-related conditions [8]. This finding is in line with the results of our study, where patients generally rated the program content as useful, but in some cases, one or two modules were rated as partly useful or not useful.

Table 4. Positive and negative effects of the C-VR and LC-VR programs on patients

	Barrier / Facilitator	ID	Quotation
C-VR			
Not forced to think critically about dose	Barrier	26, PR	I think that you could look more critically at whether it's always necessary to have 100 hours, does someone really need those 15 weeks or could they stop sooner, could more re-integration take place sooner?
Program is too uniform	Barrier	22, PR	At present, it's all very standard. I think that at the least, we could look at making the content more tailored, and get away from the kind of one-size-fits-all that we have now.
	Barrier	8, MA	At present, it's really a hit-and-miss approach and I'd want to use interventions in a more targeted way.
Redundant care	Barrier	2, MA	At the same time, we also see clients who we're currently putting in the full program, of whom we say, actually, a little less would also have been fine.
Therapy dependency	Barrier	21, PR	What we do see in the longer program, or in any case the normal one [b], is that people do build up a certain degree of dependence on the guidance, on the therapy and all.
Tailored care	Facilitator	22, PR	Look, if we're not sure whether certain parts will be feasible, then I also think yes, you know, we do want to have a go, say, a certain part that might be too much for someone, we can leave that bit out, it's not the case that the program per se has to run the way it was conceived, we can make adjustments to it, and if we want to stop earlier or even keep going for a bit longer, then we have that option.
	Facilitator	489, PT	Because it is really clearly focused on the individual, personally. And yes, that the assumption is that they look at what you can do, not at what you can't do.
Rehabilitation in a group	Facilitator	605, PT	You recognize a huge number of things that in the beginning, you always thought yourself, that it was to do with you and only you feel that, but that's not the case at all. Everyone is dealing with the same problem, in fact. So that was great, being able to recognize things in other people.
LC-VR			
Time schedule (dosage)	Facilitator	694, PT	It's really nice to be able to do those exercises at home, I was shown how to do all of them and then I was able to do them all by myself, yeah, I enjoyed that, then I didn't need to spend whole days there, say, four or five hours at a time, but normally just two or three hours.
	Facilitator	29, PR	Due to having less contact time, well, I think you're more concentrated as a result, I think that's the general added value. Clients and coaches and trainers are less able to – now, how to put it – delay things for you, wait for you.

Table 4. (Continued)

	Barrier / Facilitator	ID	Quotation
Less time spent absent from work	Facilitator	461, PT	There came a time when, more like the rest, I was already at work more and was also coming up against things, and in that way, yeah, you could share that with the group, also with the people from the [name of the rehabilitation center], in order to look at how best to deal with things if you found yourself in that kind of situation.
	Facilitator	28, PR	People have a bit more time to reintegrate, so you can make more time and space for that.
Prevention of therapy dependency	Facilitator	29, PR	... and also, that you simply empower people that way, yeah, that you can simply keep living your own life and you also have to keep on doing sports as normal, you establish a framework in that way. That they have to do it themselves. Less dependence is created.
Increase self-management	Facilitator	461, PT	I also learned to still do quite a lot myself.
	Facilitator	26, PR	Well, what I also found useful was that the client takes charge of doing things at an earlier stage, which means that we're spoon-feeding them a bit less.
Tailored care	Facilitator	313, PT	They looked specifically at what would suit me in terms of group training, because I didn't have to take part in everything. So, I did find that positive, because why should you take part in things that might not be suitable or meant for you? That might be a waste of your time.
	Facilitator	22, PR	The advantages were that it's a shorter program that's much more tailored to the individual, from the Quickscan [a] you're looking at what the person needs and how we're going to do that.
Rehabilitation in a group	Facilitator	314, PT	I think that it's very good that it's in a group and I was lucky that there were two girls of my age, who I could get along with very well. I think that all ensures that, yeah, we supported each other a lot and, you know, if someone was having a bad day, the others cheered them up, and that was really nice.

C-VR, comprehensive vocational rehabilitation; LC-VR, less comprehensive vocational rehabilitation; PT, patient; PR, professional; MA, manager.
[a] Quickscan is the center name for the multidisciplinary screening
[b] C-VR was meant here

Regarding dosage, there was a discrepancy between the opinions of patients and those of professionals. The patients were positive about the dosage of their program regardless of the actual dosage (C-VR or LC-VR). The professionals were positive about the dosage of the C-VR program and generally negative about the dosage of the LC-VR program. We assume that this discrepancy stems from the fact that the patients had no experience of VR before starting their program, whereas professionals were aware of both programs and may have been biased in favor of the C-VR program. This latter finding was also observed in another study, which found that the dosage of pain rehabilitation programs executed in clinical practice was mainly based on historical grounds and clinical experience, and not on evidence [31].

Feasibility

The patients were satisfied with the program they had been allocated (LC-VR or C-VR) and considered participating in the program to be feasible. The professionals, on the one hand, were satisfied with the C-VR program, although they did not like its fixed and uniform ("one size fits all") character and wanted more flexibility, both in terms of the content and the dosage of the program. On the other hand, the professionals had mixed views on the LC-VR program. The main argument made by professionals who had negative experiences with the LC-VR program was that it did not provide enough time to change the behavior of patients. Over the last decade, however, many RCTs [13, 14, 16] and systematic reviews [7, 10-12] have shown that the dosage of VR programs is independent of treatment outcomes (i.e., RTW). Thus, according to the present study, clinical practitioners are insufficiently aware of this finding. The managers expressed positive intentions to implement the LC-VR program in their centers (alongside the C-VR program). However, all of the managers stated that it would not be financially feasible to implement the LC-VR program, due to the structure of the Dutch healthcare system.

Strengths and limitations

By including three groups of key stakeholders, we were able to study a complex intervention such as VR from a number of different perspectives [18]. The roles (RTW coordinator, psychologist, physical therapist) of the interviewed professionals were evenly spread, which enriched the results. Of the patients who participated, ~31% were males and ~69% were females, which reflects "real world" clinical practice and thus offers a good representation of the population.

Furthermore, the study achieved a high response rate (94%) for the interviews. A further strength of the present study was that the interviews were conducted with stakeholders who had real experience of the programs of interest, enabling our findings to be transferred effectively to clinical practice.

There are also several limitations to the present study. The first is that patients who were allocated to the LC-VR program rehabilitated in the same group as patients rehabilitating in the "care-as-usual" C-VR program (and who were not included in the RCT). For financial reasons, it was not possible to create separate groups of LC-VR and C-VR patients. This flaw in the design of the RCT may have negatively influenced the experiences of patients and professionals participating in the present study. A second limitation relates to the limited experience of the professionals with the LC-VR program, which in turn limited their ability to reflect on the program. A further limitation is that recall bias may have occurred, as the period of time between the interviews and completion of the VR program was on average twelve months (patients) and six months (professionals and managers). However, another qualitative research study of the support needs of survivors of critical care found no difference in the stories of patients who underwent critical care up to five years previously [35]. This would suggest that our findings are reliable. Finally, our study was conducted in the Netherlands and therefore framed by the Dutch sickness compensation and healthcare system. We presume, however, that our findings are also representative of contexts beyond the Dutch system.

Clinical implications

The results of this study indicate that multidisciplinary VR programs could be group-based and could consist, at a minimum, of RTW coordination (communication part) and fitness sessions. Group-based education could be provided in the first weeks of the program. Other content, such as CBT, RTW coordination (ergonomic part), and relaxation sessions could be delivered to patients on a tailor-made basis. Taking the findings of the present study as a whole, we would consider it advisable to conduct quasi-flexible VR on a tailor-made basis. In order to put this into practice, we propose the following three steps: Step 1. Differentiate between C-VR and LC-VR. The patient factors proposed in the present study might assist when making this choice. Step 2. Professionals should choose from three or four blueprint programs. Step 3. Execute the program and evaluate the program together with the patient at fixed time-points (for example, after four and eight weeks). At these evaluation

moments, the decision can be made to continue with or change the content and/or dose. A final clinical implication is that key stakeholders, such as professionals, managers, and referrers, should be given clear information about the evidence underpinning a new program. In addition to all of the proposed clinical implications, however, it is of paramount importance that sickness compensation and healthcare systems facilitate the proposed changes and resources. Unless this is the case, such changes will not be feasible.

Conclusion

The patients found both programs to be feasible and generally useful. The professionals preferred working with the C-VR, but they disliked the fixed and uniform character of the program. They also mentioned that this program was too extensive for some patients, and that the latter would probably benefit from the LC-VR program. Despite their positive intentions, the managers stated that due to the Dutch healthcare system, it would not be financially feasible to implement the LC-VR program. The main conclusion of this study is that it is not useful to have one specific VR program for all patients with CMP and reduced work participation, and that quasi-flexible and tailored-based VR would thus be warranted.

Acknowledgements

This work was financially supported by Heliomare Rehabilitation Center, Wijk aan Zee, The Netherlands.

Declaration of Interest

The authors report no conflicts of interest.

Data availability

The data that support the findings of this study are available from the corresponding author, MFR, upon reasonable request.

Ethical approval

All procedures performed were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The Medical Ethical Committee of the Academic Medical Center, Amsterdam, the Netherlands, authorized this study and decided that a full application was not required.

Informed consent

Informed consent was obtained from all individual participants included in the study.

References

1. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain. 2006 May;10(4):287-333.
2. de Vroome EM, Uegaki K, van der Ploeg CP, Treutlein DB, Steenbeek R, de Weerd M, et al. Burden of Sickness Absence Due to Chronic Disease in the Dutch Workforce from 2007 to 2011. J Occup Rehabil. 2015 Dec;25(4):675-84.
3. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, et al. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. Eur Spine J. 2006 Mar;15 Suppl 2:S192-300.
4. Waddell G, Burton A, Kendall N. Vocational rehabilitation: what works, for whom, and when? The Stationery Office 2013.
5. Escorpizo R, Reneman MF, Ekholm J, Fritz J, Krupa T, Marnetoft SU, et al. A conceptual definition of vocational rehabilitation based on the ICF: building a shared global model. J Occup Rehabil. 2011 Jun;21(2):126-33.
6. van Vilsteren M, van Oostrom SH, de Vet HC, Franche RL, Boot CR, Anema JR. Workplace interventions to prevent work disability in workers on sick leave. Cochrane Database Syst Rev. 2015 Oct 5(10):CD006955.
7. Palmer KT, Harris EC, Linaker C, Barker M, Lawrence W, Cooper C, et al. Effectiveness of community- and workplace-based interventions to manage musculoskeletal-related sickness absence and job loss: a systematic review. Rheumatology (Oxford). 2012 Feb;51(2):230-42.
8. Cullen KL, Irvin E, Collie A, Clay F, Gensby U, Jennings PA, et al. Effectiveness of Workplace Interventions in Return-to-Work for Musculoskeletal, Pain-Related and Mental Health Conditions: An Update of the Evidence and Messages for Practitioners. J Occup Rehabil. 2018 Mar;28(1):1-15.
9. Waterschoot FP, Dijkstra PU, Geertzen JH, Reneman MF. Dose or content? Effectiveness of pain rehabilitation programs for patients with chronic low back pain: A systematic review. Pain 2014 Jan;155(1):179-89.
10. Hoefsmit N, Houkes I, Nijhuis FJ. Intervention characteristics that facilitate return to work after sickness absence: a systematic literature review. J Occup Rehabil. 2012 Dec;22(4):462-77.
11. Meijer EM, Sluiter JK, Frings-Dresen MH. Evaluation of effective return-to-work treatment programs for sick-listed patients with non-specific musculoskeletal complaints: a systematic review. Int Arch Occup Environ Health. 2005 Aug;78(7):523-32.
12. Schaafsma F, Schonstein E, Whelan KM, Ulvestad E, Kenny DT, Verbeek JH. Physical conditioning programs for improving work outcomes in workers with back pain. Cochrane Database Syst Rev. 2010 Jan 20(1):CD001822.
13. Aasdahl L, Pape K, Vasseljen O, Johnsen R, Gismervik S, Halsteinli V, et al. Effect of Inpatient Multicomponent Occupational Rehabilitation Versus Less Comprehensive Outpatient Rehabilitation on Sickness Absence in Persons with Musculoskeletal- or Mental Health Disorders: A Randomized Clinical Trial. J Occup Rehabil. 2017 Apr 11.

14. Ronzi Y, Roche-Leboucher G, Begue C, Dubus V, Bontoux L, Roquelaure Y, et al. Efficiency of three treatment strategies on occupational and quality of life impairments for chronic low back pain patients: is the multidisciplinary approach the key feature to success? Clin Rehabil. 2017 Oct;31(10):1364-73.

15. Haldorsen EM, Grasdal AL, Skouen JS, Risa AE, Kronholm K, Ursin H. Is there a right treatment for a particular patient group? Comparison of ordinary treatment, light multidisciplinary treatment, and extensive multidisciplinary treatment for long-term sick-listed employees with musculoskeletal pain. Pain. 2002 Jan;95(1-2):49-63.

16. Skouen JS, Grasdal AL, Haldorsen EM, Ursin H. Relative cost-effectiveness of extensive and light multidisciplinary treatment programs versus treatment as usual for patients with chronic low back pain on long-term sick leave: randomized controlled study. Spine (Phila Pa 1976). 2002 May 1;27(9):901-9.

17. Loisel P, Buchbinder R, Hazard R, Keller R, Scheel I, van Tulder M, et al. Prevention of work disability due to musculoskeletal disorders: the challenge of implementing evidence. J Occup Rehabil. 2005 Dec;15(4):507-24.

18. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ. 2008 Sep 29;337:a1655.

19. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med. 2009 May;36(5):452-7.

20. Williams-Whitt K, Bultmann U, Amick B, 3rd, Munir F, Tveito TH, Anema JR. Workplace Interventions to Prevent Disability from Both the Scientific and Practice Perspectives: A Comparison of Scientific Literature, Grey Literature and Stakeholder Observations. J Occup Rehabil. 2016 Dec;26(4):417-33.

21. Wind H, Gouttebarge V, Kuijer PP, Sluiter JK, Frings-Dresen MH. The utility of functional capacity evaluation: the opinion of physicians and other experts in the field of return to work and disability claims. Int Arch Occup Environ Health. 2006 Jun;79(6):528-34.

22. Abbott JH. The distinction between randomized clinical trials (RCTs) and preliminary feasibility and pilot studies: what they are and are not. J Orthop Sports Phys Ther. 2014 Aug;44(8):555-8.

23. Beemster TT, van Velzen JM, van Bennekom CA, Frings-Dresen MH, Reneman MF. Cost-effectiveness of 40-hour versus 100-hour vocational rehabilitation on work participation for workers on sick leave due to subacute or chronic musculoskeletal pain: study protocol for a randomized controlled trial. Trials. 2015 Jul 28;16:317.

24. 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 Dec;19(6):349-57.

25. Wet verbetering poortwachter (The Law on Gate keeping Disability Insurance). Accessed 22 Jan 2018; Available from: https://www.arboportaal.nl/onderwerpen/wet-verbetering-poortwachter.

26. Fleuren MA, Paulussen TG, Van Dommelen P, Van Buuren S. Towards a measurement instrument for determinants of innovations. Int J Qual Health Care. 2014 Oct;26(5):501-10.

27. Flottorp SA, Oxman AD, Krause J, Musila NR, Wensing M, Godycki-Cwirko M, et al. A checklist for identifying determinants of practice: a systematic review and synthesis of frameworks and taxonomies of factors that prevent or enable improvements in healthcare professional practice. Implement Sci. 2013 Mar 23;8:35.

28. Research Code VUmc-AMC. Accessed 22 Jan 2018; Available from: https://www.vumc.nl/onderzoek/research-code/.

29. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology. 2006;3(2):77-101.

30. Waterschoot FP, Bennen E, van der Woude LH, Schiphorst Preuper HR, Reneman MF. Case complexity in patients with chronic nonspecific musculoskeletal pain: a Delphi and feasibility study. Int J Rehabil Res. 2016 Mar;39(1):48-56.

31. Reneman MF, Waterschoot FPC, Bennen E, Schiphorst Preuper HR, Dijkstra PU, Geertzen JHB. Dosage of pain rehabilitation programs: a qualitative study from patient and professionals' perspectives. BMC Musculoskelet Disord. 2018 Jun 30;19(1):206.

32. Loeb DF, Binswanger IA, Candrian C, Bayliss EA. Primary care physician insights into a typology of the complex patient in primary care. Ann Fam Med. 2015 Sep;13(5):451-5.

33. Selander J, Marnetoft SU, Bergroth A, Ekholm J. Return to work following vocational rehabilitation for neck, back and shoulder problems: risk factors reviewed. Disabil Rehabil. 2002 Sep 20;24(14):704-12.

34. Viikari-Juntura E, Virta LJ, Kausto J, Autti-Ramo I, Martimo KP, Laaksonen M, et al. Legislative change enabling use of early part-time sick leave enhanced return to work and work participation in Finland. Scand J Work Environ Health. 2017 Sep 1;43(5):447-56.

35. Allum L, Connolly B, McKeown E. Meeting the needs of critical care patients after discharge home: a qualitative exploratory study of patient perspectives. Nurs Crit Care. 2017 Jun 22.

Appendix 1. Topic lists for patients, professionals, and managers concerning the usefulness and feasibility of C-VR and LC-VR

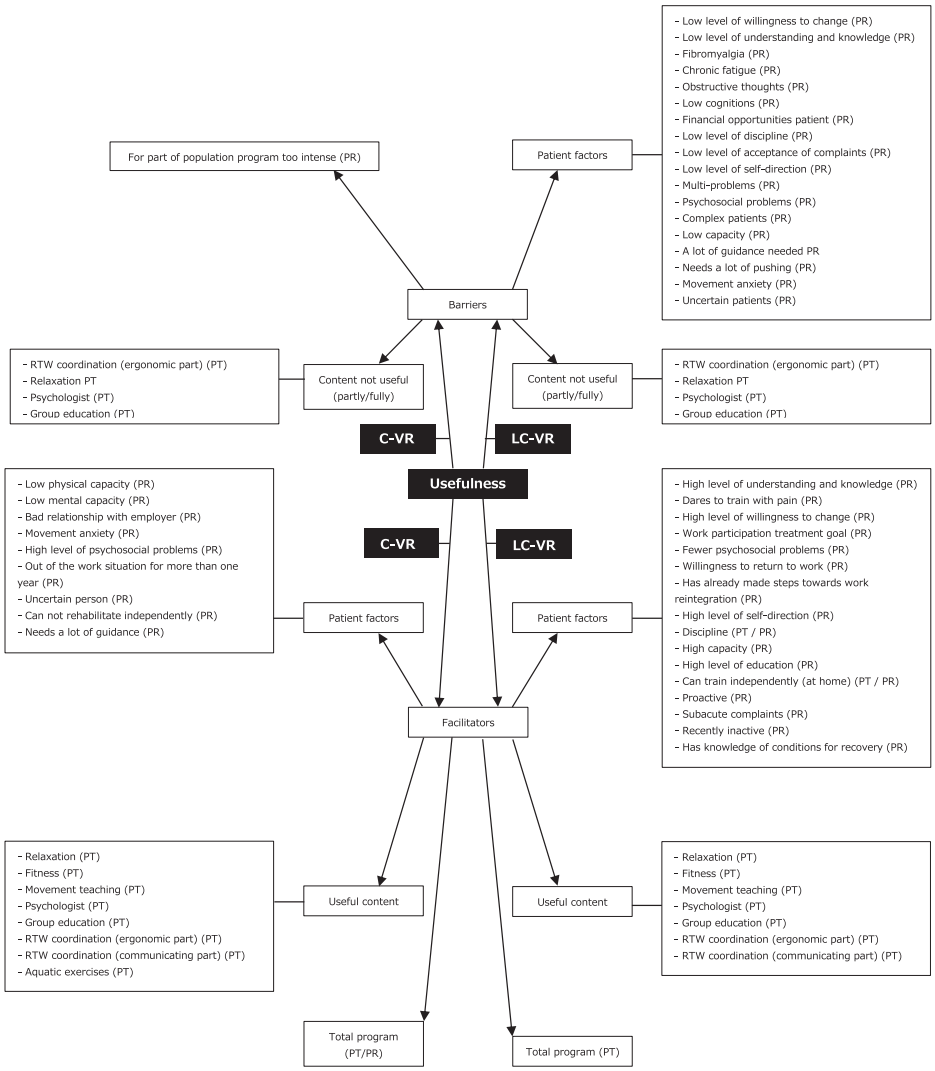
		Program (C-VR, LC-VR)
	Patients	
Feasibility	· Satisfaction with the program (on a 0-10 scale: 0=not satisfied at all, 10=very satisfied). · Goals achieved? Work functioning improved? · Feasibility of program in terms of time management. · Recommend the program to others?	C-VR, LC-VR
Dosage	· Dosage of the program: contact hours, frequency, and program duration.	C-VR, LC-VR
Usefulness	· Usefulness of content[a]: fitness, sessions with psychologist, RTW coordination (ergonomic part and communication part), group education, relaxation sessions, or other modules. · Content most/less useful. · Rehabilitation in a group.	C-VR, LC-VR
Adaptations	· Adaptations to improve the program.	C-VR, LC-VR
	Professionals	C-VR, LC-VR
Usefulness	· Useful for the patient? · Suitability of the program for all patients? · For which patients was the program useful/would the program be useful? · For which patients was the program not useful/would the program not be useful?	C-VR, LC-VR
Feasibility	· Feasible of continuing (C-VR) / implementing (LC-VR) in the future? · Benefits of the program compared with the other[b] program.	C-VR, LC-VR
Adaptation	· Adaptations to make the program more suitable for patients? · Adaptations to improve the program?	C-VR, LC-VR
Choice	· If you had to make a choice, which program would you prefer?	C-VR, LC-VR
Implementation	· Barriers to/facilitators for implementation.	LC-VR
Intention	· Willingness to work with the program in the future.	LC-VR

Appendix 1. (Continued)

		Program (C-VR, LC-VR)
	Managers	
Feasibility · If yes: · If no:	· Feasibility of replacing the current (C-VR) program with the LC-VR program in the near future? · Implementation factors (such as: financial/reimbursement factors, logistical issues, cooperation of professionals, etc.). · Anticipated satisfaction of referrers. · Implementation barriers. · How much shorter (in terms of contact hours) than the C-VR program would be feasible for your center? · Changes needed to offer a LC-VR program as a/the standard program in your center?	LC-VR
Feasibility	· Feasibility of conducting the C-VR program in the near future? · Barriers to/facilitators for continuation. · Satisfaction of referrers with the C-VR program.	C-VR

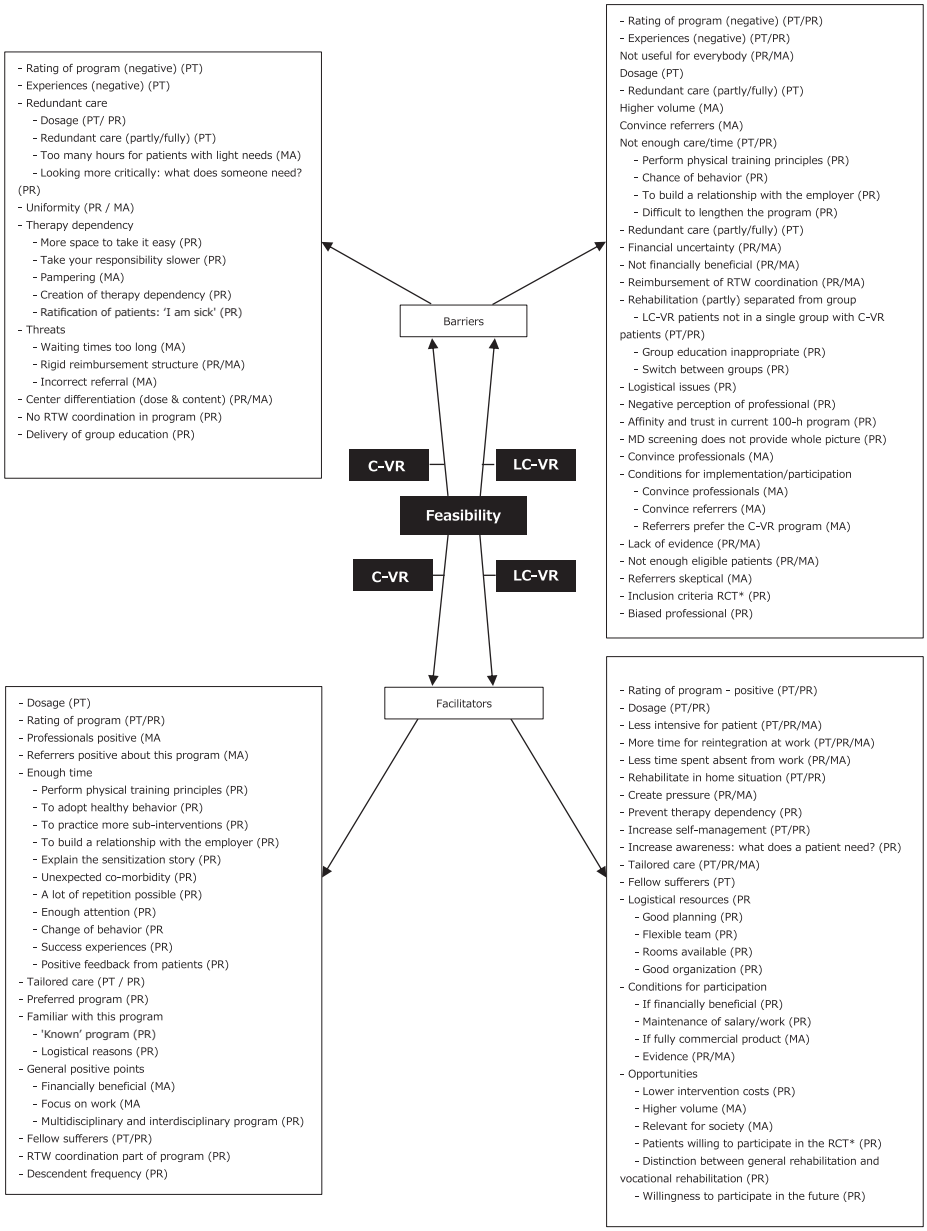
C-VR, comprehensive vocational rehabilitation; LC-VR, less comprehensive vocational rehabilitation
[a] We asked about the usefulness and dosage of each module.
[b] C-VR versus LC-VR or LC-VR versus C-VR.

Appendix 2. Code tree: usefulness

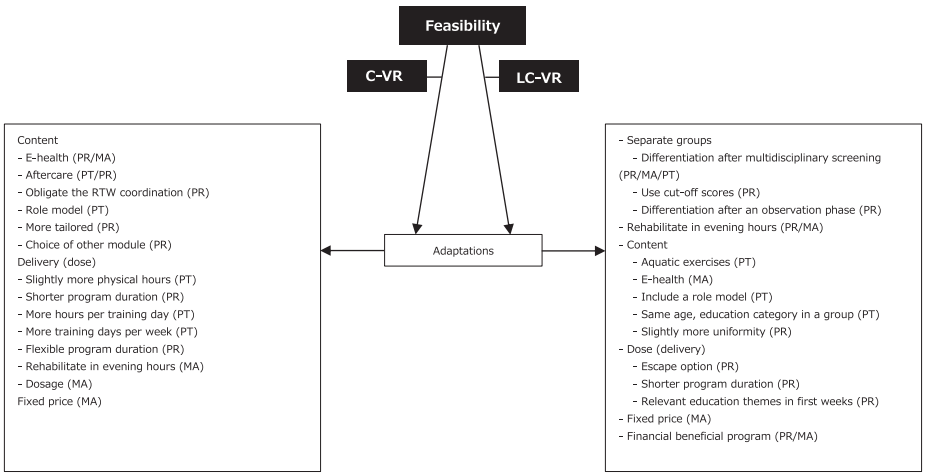


C-VR, comprehensive vocational rehabilitation; LC-VR, less comprehensive vocational rehabilitation; PT, patient; PR, professional; RTW, return to work

Appendix 3. Code tree: feasibility



Appendix 3. Code tree: feasibility (continued)



C-VR, comprehensive vocational rehabilitation; LC-VR, less comprehensive vocational rehabilitation; PT, patient; PR, professional; MA, manager; RTW, return to work; RTW, return to work; RCT, Randomized Controlled Trial
* Beemster et al. [23]

Appendix 4. Quotations of patients, professionals, and managers concerning the barriers to and facilitators of the dosage of C-VR and LC-VR

Codes	Barrier / Facilitator	ID	Quotation
C-VR			
Hours per day	Barrier	598, PT	I would have liked to have been able to keep going for longer (...) certainly a half-day, yes.
	Facilitator	522, PT	It was doable, but it really was no longer necessary. [In relation to the number of hours training per day]
Program duration	Barrier	522, PT	It would have been fantastic, yes, if it could have gone on for a bit longer.
	Facilitator	489, PT	But I was satisfied with how it went, say, 15 weeks was just right.
	Barrier	598, PT	Perhaps a little more training, all the same, or more training, several times a week. (...) And that the whole course could then be shorter.
Dosage/frequency	Facilitator	605, PT	For me, the amount of time was actually precisely enough, it was good as it was.
	Barrier	476, PT	I think that it was more than enough in the first week, and in my opinion, they could have extended the final week by a day or half-day.
The C-VR program could be slightly shorter	Barrier	21, PR	I think that all of us, in principle, could have achieved the same result in fewer hours, I do think that, but how many hours, I don't know, but I do think less.
	Barrier	26, PR	That period of 15 weeks is, in my opinion, also a bit arbitrary, perhaps it could be done in 12 weeks.
	Barrier	29, PR	I can imagine that it might be possible to reduce it to around 10 weeks, and you could perhaps set more assignments digitally.
Less complex patients	Barrier / Facilitator	26, PR	Because within the 100 hours, you sometimes notice that people need more time for a behavioral change so that they can pick things up, so I expect that for a certain group it would indeed be OK, but there are also people who, well, now, need more intensive guidance.
	Barrier	2, MA	At the same time, we also see clients who we're currently putting in the full program, of whom we say, actually, a little less would also have been fine.
Tailored care	Facilitator	28, PR	If I had to state a preference, I would say, goodness, I would go for the 100-hour program, because you can always tailor it.

Appendix 4. (Continued)

Codes	Barrier / Facilitator	ID	Quotation
	Facilitator	22, PR	Look, if we're not sure whether certain parts will be feasible, then I also think yes, you know, we want to have a go, say, but a certain part might be too much for someone, so we can leave that bit out, it's not the case that the program per se has to run the way it was conceived, we can make adjustments to it, and if we want to stop earlier or even keep going for a bit longer, then we have that option.
Enough time:	Facilitator	25, PR	You have a better overview of someone, otherwise it's a very short time to get to know someone really well, learn how they think and what their personality is like, how they cope, you need a bit more time for that, otherwise I think it would become a very physical story and the rest, well, there'd simply be little time for it.
• Perform physical training principles	Facilitator	30, PR	In the physical training, you go through a number of phases, and if you had 40 hours, you wouldn't get around to all those training phases, and in 100 hours you would, meaning that in the end, the physical resilience simply becomes greater if you're working longer and more intensively.
• Change of behavior	Facilitator	30, PR	Because you have enough time to bring about the change in behavior which is a big part of the program.
	Facilitator	30, PR	But we see that the physical and the mental go hand in hand, of course, such as physically non-specific lower back pain, if someone has been suffering that for six months, then you also have to get the sensitization story out of the brain, that's something you actually have to build up step-by-step in the training, so it's not the case that if someone doesn't have a disorder, they'll be able to finish sooner. It really is a combination with mental and behavioral.
• Adopt healthy behavior	Facilitator	29, PR	That they pick up behavior, such as working more, doing more sport, moving around more, doing nice things for themselves, you have more time for that.
• Explore extra interventions	Facilitator	29, PR	(...) it means you can practice more interventions, such as communication tasks.
• Unforeseen co-morbidity	Facilitator	25, PR	Sometimes things do happen during the course, someone might get a diagnosis that wasn't known about beforehand, including in the Quickscan [multidisciplinary screening].
• Build a relationship with the employer	Facilitator	29, PR	(...) let people tell their employers what they consider to be important in re-integration.

Appendix 4. (Continued)

Codes	Barrier / Facilitator	ID	Quotation
LC-VR			
Hours per day	Barrier	313, PT	I mean I was already there and would have liked to work out for another hour under supervision.
	Facilitator	694, PT	It's really nice to be able to do those exercises at home, I was shown how to do all of them and then I was able to do them all by myself, yeah, I enjoyed that, then I didn't need to spend whole days there, say, four or five hours at a time, but normally two or three hours, and then I could continue at home.
Program duration	Barrier	461, PT	Yeah, what I remember of that time is that I really found it a shame that it stopped and, um, I would have liked to have been able to keep going for longer.
	Facilitator	694, PT	It really gives you a limit, you actually determine the limit, and you do figure it out at a certain point, after 10-12 weeks or so. And then it's really a question of simply ensuring that in the final weeks you do the exercises properly, so that you can do them on your own.
Frequency	Facilitator	461, PT	There were fewer hours in the final weeks, but I thought it was great that I still had that incentive every time.
	Facilitator	212, PT	Well, in fact I found those two times a week great (...) also because I was still doing exercises at home.

C-VR, comprehensive vocational rehabilitation; LC-VR, less comprehensive vocational rehabilitation; PT, patient; PR, professional; MA, manager

Appendix 5. Quotations of patients, professionals, and managers concerning the barriers and facilitators of feasibility themes of C-VR and LC-VR

Themes and codes	Barrier / Facilitator	Program	ID	Quotation
Intention to continue use				
Recommend the program to family, friends, and colleagues	Facilitator	C-VR	605, PT	Yes, I already spoke with people who were indeed in the same process or dealing with the same problems, and who talked about how you found it. And I said very clearly that it turned out very positive for me and that I would definitely recommend it.
	Facilitator	LC-VR	694, PT	Yes, definitely. I certainly got a lot out of it. Certainly, as I already said, a bit of acceptance et cetera, and the exercises I can do to relieve my back.
Willingness to work with the LC-VR program in the future	Facilitator	LC-VR	24, PR	If they tell me that I have to do something, then I do it, because I think I should do it for a boss, and otherwise I would have done my very best to set it up and make it happen, but if I can't completely support it, that's another matter, but I'll certainly do it.
	Facilitator	LC-VR	29, PR	This intention is positive for the matching target group, as I sketched out for you earlier.
Perceived appropriateness				
C-VR program as the most appropriate program:	Facilitator	C-VR	21, PR	I can't choose between the two. I would prefer in between the two, but if I had to choose, I would go for 100 hours.
• Enough time/ tailored care	Facilitator	C-VR	28, PR	If I had to state a preference, I would say, goodness, I would go for the 100-hour program, because you can always tailor it.
• Current program	Facilitator	C-VR	24, PR	Now we have a program, I think, that is very well matched to what we want.
• 'Known' program	Facilitator	C-VR	22, PR	(...) I would probably choose the current program and that's particularly to do with, well, yes, who I am as a person, I tend to prefer to take the familiar path, it's easier to fall back on that.
• Logistical reasons	Facilitator	C-VR	22, PR	From an organizational point of view, it is really a lot easier, because it's simply a fixed program for all of the convalescents who are in the group.
• Financially beneficial	Facilitator	C-VR	5, MA	For us, it's one of the things that keeps us afloat. (...) In that sense, it's also about continuity of management.

Appendix 5. (Continued)

Themes and codes	Barrier / Facilitator	Program	ID	Quotation
Factors affecting ease or difficulty of implementation				
Too much diversity	Barrier	LC-VR	21, PR	If you only had 40-hour programs and you looked at everyone individually to see what their needs are, I think that would become a pretty tricky puzzle.
	Barrier	LC-VR	26, PR	You also try to balance what is the best possible for the client against what is organizationally feasible, that's where the tension lies.
LC-VR and C-VR in one group	Barrier	LC-VR	28, PR	For us, in practice, the 40-hours took part with the 100 hours. I found that this created a bit of a mess.
	Barrier	LC-VR	2, MA	You shouldn't have people who are doing a 40-hour program and people who are doing a 100-hour program in a single group, because that gives a strange impression to clients who talk to and influence each other.
Lack of evidence	Barrier	LC-VR	26, PR	But actually, you would like to have a better foundation of, yes, who it's more and less suitable for.
Best practices	Barrier	LC-VR	3, MA	Naturally, I don't have any reason, scientifically or whatever, or in the area of best practice, to put a stop to a load of existing practices and transform them into a program of a single ... size.
Prejudiced	Barrier	LC-VR	26, PR	On the basis of the intake, as a team, you have a certain idea of what someone needs and, well, perhaps you have a tendency, if it seems a bit more complex, to say that someone needs more. Anyway, when someone wants to take part in the trial and takes part in the 40 hour-program, then we've also seen a good few times that they've actually made quite good progress.
Reimbursement of program	Barrier	LC-VR	1, MA	It's the [Dutch] reimbursement structure itself that is a potential problem, that you'd have to decide, unfortunately, that that makes it financially unfeasible.
	Facilitator	LC-VR	22, PR	The most important thing is the financing, if that works well, then I think that it's feasible.

Appendix 5. (Continued)

Themes and codes	Barrier / Facilitator	Program	ID	Quotation
Reimbursement of RTW coordination	Barrier	Both [a]	22, PR	Managing to ensure that the reintegration guidance is also paid for by the employer, and that's where we still see, yes that's still the most difficult, finding employers who are ready to pay for it. They say, yes, we're dealing with that ourselves, and yes, that's always difficult.
Rigid financial structure	Barrier	Both	1, MA	And if it's shown that it really is much more effective than doing more hours, then it is strange that we in the Netherlands have a funding structure that makes that more likely, I think that's a bit of the worry that we'll be doing yet more hours. We know that it doesn't make any sense, but then we do get slightly higher rates. Yes, that's of course really stupid.
	Barrier	Both	5, MA	I have to earn money is order to keep my people here in work. Well, that can be achieved by putting some additional time into a patient's treatment.

C-VR, comprehensive vocational rehabilitation; LC-VR, less comprehensive vocational rehabilitation; PT, patient; PR, professional; MA, manager. [a] Implementation factor for both programs (i.e., C-VR and LC-VR).

Appendix 6. Recommended adaptations of the content and delivery of C-VR and LC-VR, according to patients, professionals, and managers

Codes	Program	ID	Quotation
Content			
Slightly more uniform	LC-VR	22, PR	There you'd probably look for something more, yeah, not that one-size-fits-all, but, say, something in between, that can be set up well, especially in a scheduling, organizational sense (...) you could maybe work towards having a few standard forms, two or a maximum of three, where you could rehabilitate people if you took the structure, if you took the planning into account, then it would probably be more manageable.
RTW coordination module obligatory	C-VR	29, PR	I think that the contribution of the work coaches [RTW coordinator], um, is actually quite important in this. Because they often start with a build-up in the program, of working hours, and I think that's a crucial, yes (...) condition for work-coaching [RTW coordination].
More tailored	C-VR	22, PR	If you look at the different components that come up in the rehabilitation program, you should look more at what someone needs and how often someone needs it.
	C-VR	8, MA	I think it's desirable to, say, take a more targeted approach, use more targeted resources, I think that at the moment, in any case, we really take a hit-and-miss approach, and I would want to use interventions in a more targeted way.
More choice of content	Both [a]	28, PR	That you work more with modules.
Aftercare	C-VR	598, PT	What's missing for me is an evaluation day.
	C-VR	25, PR	We are actually planning to introduce an evaluation day, we do also think that's important.
	LC-VR	314, PT	I think that it would be nice if you were to get one more email from the psychologist or from the RTW coordinator, saying, how's it going, asking whether they could do anything else for you. I think that wouldn't do any harm.
Role model	C-VR	605, PT	Perhaps it would be an idea, if you're working on that kind of program, for the practitioner to maybe involve a kind of practical expert in the program, so that people who've done the program were asked to come back and say something about it, perhaps?
E-health	LC-VR	5, MA	That more is done at home with the participants, for example, say, e-health is also a resource that you should be able to use in order to have the participants do things at home outside patient time at the rehabilitation center.
	C-VR	29, PR	Read tasks via email or e-health and then do them, um, but then do them in their own living environment, and I think that would be much more effective than having them stuck in a kind of setting.
Swimming	LC-VR	694, PT	I don't know if it's possible, but I have really benefited from swimming.

Appendix 6. (Continued)

Codes	Program	ID	Quotation
Delivery			
Reduce program duration	LC-VR	30, PR	So, somewhat milder mental, somewhat milder physical symptoms, for which you think that you can, yeah, also achieve a great deal in 8-10 weeks.
	C-VR	28, PR	Someone might also be able to come a long way in 12 weeks.
Flexible program duration	C-VR	22, PR	What we're thinking about now, in any case, is do we want to offer 15 weeks per se, or shall we leave it more open, so that after some time we think about how far someone has come and look at whether they have already achieved their goals, or where they have come from and how long you will keep going, that you're also more flexible about this.
Reduce frequency over time	C-VR	22, PR	For us, the pattern is that from beginning to end, you come twice a week, and you could also cut that back over time, for example, so you could say that for so many weeks, you come twice a week, and perhaps at the end you come just once a week, and do the other day something for yourself, at least that's an approach that I think certainly wouldn't do any harm.
Educational themes	LC-VR	694, PT	I found it useful, but at a certain point, it all became much of a muchness, if you know what I mean. At a certain point, you know what kind of pain Peter has and what kind of pain Paul has.
	C-VR	26, PR	You have a number of themes, for example, for education, yeah, ideally you would put certain themes at the beginning of the program as group meetings, and other ones later, but that is a bit inherent in the open groups that we're running.
Separate programs	Both	24, PR	The way it is now, that you have people doing the 40-hour program in the same space as those doing the 100-hour one, and those doing the 100 go swimming and you have to say to those doing the 40, sorry, that's not possible in your program, that feels really stupid, so you should actually have separate groups, I think.
	Both	2, MA	You shouldn't have people who are doing a 40-hour program and people who are doing a 100-hour program in a single group, because that gives a strange impression to clients who talk to and influence each other.
Differentiation between C-VR and LC-VR	Both	21, PR	You're always extremely limited in what you can see in a Quickscan [multidisciplinary screening] like that, your picture is enlarged because there are all kinds of perspectives and different people, but ultimately you actually see much more during that kind of course, because then you see someone in different situations and more often.
• Cut-off scores	Both	26, PR	I'm thinking of certain, in any case certain cut-off scores, so if you say from, well, above that much, yeah, above a certain score, maybe it's less indicated.

Appendix 6. (Continued)

Codes	Program	ID	Quotation
• Observation phase	Both	26, PR	Within pain rehabilitation, for example, for some time they had an observation phase of three weeks, and if I think about that now, then I think, hey, you can also see a bit whether people pick things up, so you might able to do something like that, an observation phase lasting a few weeks and then decide whether it's going to be 40 or 100 hours.
• Same age/ education category in one group	Both	314, PT	If possible, you could try to put younger people together, because in terms of age, um, it's a category that has lots in common and they also understand each other's environment better. (...) it sounds really bad, but I could get irritated by the level and by people who constantly asked the same questions in the thematic meetings, or by people who just didn't get it, and that's very frustrating for them, and it's really not that I think those people are stupid, but after a while, it can get quite annoying.
Group size	Both	29, PR	I think that a group of eight, say, is a good number, and if they are clients with minor needs, then ten wouldn't be a problem either, that's something you need as a condition for the group dynamic.
Escape option	LC-VR	21, PR	Only it can get stuck there, when we see that we can't manage them enough with 40 hours, that's where it gets stuck.

C-VR, comprehensive vocational rehabilitation; LC-VR, less comprehensive vocational rehabilitation; PT, patient; PR, professional; MA, manager. [a] Adaptation for both programs (i.e., C-VR and LC-VR).

CHAPTER 7

Vocational Rehabilitation with or without Work Module for Patients with Chronic Musculoskeletal Pain and Sick Leave from Work: Impact on Work Participation

T. Beemster
C. van Bennekom
J. van Velzen
M. Frings-Dresen
M. Reneman

Submitted

Abstract

Purpose

To study the relationship between interdisciplinary vocational rehabilitation with (VR+ program) or without (VR program) additional work module on work participation of patients with chronic musculoskeletal pain and sick leave from work.

Methods

A retrospective cohort study was conducted, with data retrieved from care as usual in seven VR centers in the Netherlands. The VR program consisted of multi-component healthcare (physical exercise, cognitive behavioral therapy, education, relaxation). VR+ additional components were case management and workplace visit. The dependent variable was work participation (achieved/not achieved). Independent variables were type of intervention (VR/VR+), demographics, clinical, and work-related (return to work [RTW] expectation, sick leave duration, working status, job strain, and job dissatisfaction). Multivariate logistic regression analyses were applied on discharge and six-months follow-up.

Results

Of the 142 patients included, 26% received VR and 74% VR+. Both programs increased work participation at six-months follow-up (VR 80%, VR+ 86%). There were non-significant relationships between type of intervention and work participation on discharge (OR 1.0, $p = 0.99$) and six-months follow-up (OR 1.3, $p = 0.52$). RTW expectation was the only significant independent factor in the multivariate model on discharge (OR 2.9, $p = 0.00$) and six-months follow-up (OR 3.0, $p = 0.00$).

Conclusions

Both programs led to increased work participation. The addition of a work module to the VR program did not lead to significant increase in odds of work participation at discharge and six-months follow-up. This finding was probably due to a lack of contrast between the two programs.

Keywords

Chronic pain, observational study, occupational therapy, biopsychosocial, multidisciplinary.

Introduction

Chronic musculoskeletal pain (CMP) affects quality of life, disability, and work [1, 2]. Workers with CMP have high rates of absenteeism and presenteeism (at work but with decreased productivity), with productivity losses equivalent to 1.6% of Gross Domestic Product for the Netherlands [3]. Thus, the main goal of interventions for patients with CMP and productivity loss from work is to increase work participation. Several reviews have shown that interdisciplinary vocational rehabilitation (VR) programs are effective in realizing this goal [4-6].

There is large variation in the content of VR programs [4-7]. A recent review recommended that effective VR programs should encompass the following three domains: 1. health-focused (i.e., health services intervention subcategories such as graded activity/exercise, cognitive behavioral therapy [CBT], work-hardening), 2. service coordination (i.e., improving communication within the workplace or between the workplace and the healthcare providers), and 3. work modification (i.e., modified duties, modified working hours, supernumerary replacements, ergonomic adjustments, or other worksite adjustments) (Box 1) [4]. The same review also mentioned that a multi-domain intervention including components in at least two of the three domains mentioned, can help reduce lost time from work for CMP-related conditions [4].

The review mentioned above and other studies on this topic mainly consist of RCT studies in which multi-domain programs were compared with usual care [4, 5, 8] or with single component programs from the health-focused domain, such as graded activity/physical exercise [5, 7, 9], or education [5]. Little evidence is available about the additional increase in effect on work participation when components from the work-related domains (i.e., service coordination and work modifications, see Box 1) are added to a multi-component health-focused program. The latter is standard care for patients with CMP in most industrialized countries. However, the evidence concerning this niche is contradictory.

On the one hand, an RCT study conducted in Norway in patients with neck and back pain found no significant differences in work participation between the group who took part in a multidisciplinary program (i.e., multi-components from the health-focused domain) that included work-focused components and a group who only took part in a multidisciplinary program [10]. On the other hand, a retrospective cohort study conducted in Canada showed that a multidisciplinary (i.e., multi-components from the health-focused domain) pain program that

included return to work coordination had 3.4 higher odds of a return to work compared with a multidisciplinary program without coordination [11].

In summary, while the evidence on the overall effectiveness of VR is robustly positive, the evidence concerning the content of VR is contradictory. In the present study, we analyzed the difference in work participation of patients who were referred to multi-component health-focused VR program with or without an additional work module in clinical practices in the Netherlands (VR+ and VR respectively).

The research question of this study was: Are patients with CMP who are on sick leave from work more likely to participate in work if they take part in a VR+ program compared with patients who only take part in a VR program? Based on recommendations from various systematic reviews to include work domains in VR to achieve successful work participation [4, 5, 8, 12], we hypothesized that patients who took part in the VR+ program would have higher odds of participating in work compared to patients who only took part in the VR program.

Box 1. Intervention components in rehabilitation treatments

Health-focused interventions. These interventions facilitate the delivery of health services to the injured worker either in the workplace or in settings linked to the workplace (e.g., visits to healthcare providers initiated by the employer/workplace). Specific health services intervention subcategories for which evidence synthesis was conducted include; graded activity/ exercise, cognitive behavioural therapy, work hardening and multi-component health-focused interventions (which often included the above elements as well as: medical assessment, physical therapy, psychological therapy, occupational therapy).

Service coordination interventions. These interventions were designed to better coordinate the delivery of, and access to, services to assist RTW within and involving the workplace. Coordination involves attempts to improve communication within the workplace or between the workplace and the healthcare providers. Examples are development of RTW plans, case management and education and training.

Work modification interventions. These interventions alter the organization of work or introduce modified working conditions. Examples are: workplace accommodations such as provision of modified duties, modified working hours, supernumerary replacements, ergonomic adjustments or other worksite adjustments.

Multi-domain interventions. These interventions had multiple intervention components and included at least two of the three above intervention domains [e.g., interventions that involved graded activity in the workplace (health-focused domain) in addition to modified working conditions (work modification domain)].

Text obtained from Cullen et al. [4]

Methods

The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist was used in the design and reporting of this study [13].

Design, setting, and procedure

A retrospective cohort study was conducted, with data collected from November 2014 to July 2018 by seven rehabilitation centers located throughout the Netherlands. These seven centers all offered interdisciplinary VR for workers with CMP who were hampered in their work participation. Patients were referred to the VR program by their occupational physician, general physician, rehabilitation physician, medical specialist, or others. Before entering the VR program, patients completed web-based questionnaires (T0) and underwent a multidisciplinary (MD) screening performed by an MD team consisting of a rehabilitation physician, psychologist, physical therapist, and vocational specialist. After the MD screening, the team and patient decided whether a VR+ program was appropriate or not (criteria, see [14]). Before VR+ started, the employer of every patient was asked to reimburse the additional work module (€1200), which was a condition of the patient participating in the VR+ program. VR was reimbursed by the healthcare insurer. Apart from the additional work module, patients of both programs participated as one group. Patients received web-based questionnaires at discharge (T1) and at six-months follow-up (T2). If patients did not complete the T0-2 questionnaires within a week, they received a reminder by email.

Participants

Working age individuals (18-65 years) with subacute or chronic musculoskeletal pain and reduced work participation (full or part-time sick leave) who were referred to vocational rehabilitation and who underwent a vocational rehabilitation program (VR+ or VR) between September 2014 and October 2017 participated in this study. Patients were excluded if they had no paid work, if they were not able to complete questionnaires in Dutch, or if they did not grant informed consent. The Medical Ethical Committee of the Academic Medical Center, Amsterdam, the Netherlands, authorized this study and decided that a full application was not required (number W18_194). Participation in the study was voluntary, all participants provided informed consent, and answers were processed anonymously.

Context

When an employee is sick-listed in the Netherlands, both the employee and employer are responsible for the work participation process during the first two years of sick leave. According to the Dutch Gatekeeper Improvement Act, the employer must provide wage replacement and modified work during this two-year period [15].

Interventions

Vocational rehabilitation (VR)
The vocational rehabilitation (VR) program was an interdisciplinary group-based program that consisted of multi-components from the health-focused domain. They included general exercise therapy based on principles of graded activity (total ~60 hours; 30 x 2 hours), CBT (total ~7.5 hours; 15 x 0.5 hour), group education (total ~15 hours; 15 x 1 hour), and relaxation (total ~7.5 hours; 15 x 0.5 hour). There were two evaluation moments with the patient: one mid-evaluation after seven weeks and one end evaluation at discharge. A report from these two evaluation moments was sent to the patient. The MD team consisted of a physician, physiotherapist, and a psychologist. The program lasted fifteen weeks (total ~90 hours) with two 3.5 to 4 hour sessions per week. More information about the content of the VR program can be found in the study protocol paper [16].

Vocational rehabilitation + work module (VR+)
The vocational rehabilitation + work module (VR+) program was an interdisciplinary group-based program that consisted of the same health-focused components as the VR program, but was extended with a work module. The work module consisted of case management and a workplace visit (total of ~10 hours), and was executed by an RTW coordinator. The case management involved discussion of work-related problems, the design and discussion of the progress of a work participation plan, and the provision of information about work-related legislation. The company visit included communication between the patient, the RTW coordinator, and the employer with the goal of discussing and resolving barriers to and facilitators of work participation, as well as discussing a work participation plan. A workplace inspection with possible advice for ergonomic adjustment was also part of the workplace visit. There were two evaluation moments with the patient: one mid-evaluation after seven weeks and one end evaluation at discharge. A report of these two evaluation moments was sent

to the patient and his/her employer and occupational physician. If necessary, the evaluation reports were discussed with the employer and/or occupational physician. The MD team consisted of a physician, physiotherapist, psychologist, and an RTW coordinator. The program lasted fifteen weeks (total ~100 hours) with two 3.5 to 4 hour sessions per week. An outline of the content and dosage of the modules of the VR+ program are described in the study protocol paper [16].

Measures

Dependent variable: work participation

Work participation was assessed using the *working status* item of the *imta Productivity Cost Questionnaire-Vocational Rehabilitation version* (iPCQ-VR) [17]. Working status was assessed with the question: "Are you working full-time at this moment?" with the answer categories: "Yes," "No, I am partly at work," and "No, I am on 100% sick leave." In the case of patients being partly at work, there was an additional question: "How many hours are you working per week at the moment?" For the aim of this study, the *working status* and *hours working per week* items were first converted into a continuous variable of "hours working per week." In a second step, the change in working hours per week was calculated by subtracting working hours per week at T1/T2 from the working hours per week at T0. In a final step, the working hours per week difference was dichotomized into "Achieved work participation" for those who worked at least one hour or more per week at T1/T2 compared to T0, and "Not achieved work participation" for those who worked the same working hours per week or less at T1/T2.

Independent variables

The fixed independent variable in this study was **type of intervention** (VR+/VR). The other independent variables selected were potentially associated with or confounders of the outcome of "work participation." The independent variables of this study were clustered into biopsychosocial characteristics [18]: demographic, personality, disorder-related, and work-related. Hereafter, we briefly describe the content and score ranges of the independent variables selected and used in this study. A detailed description and clinometric properties of the questionnaires included can be found elsewhere [16, 17, 19].

Demographic characteristics

The following demographic characteristics were included: **age** [20-23], **gender** [11, 21-24], and **level of education** [22, 25-28]. Age was dichotomized based on the median. Level of education was divided into three categories: "low" (including primary school, lower vocational education, and lower secondary school), "medium" (including intermediate vocational education and upper secondary school), and "high" (including upper vocational education or university) [25].

Psychological variables

The following psychological characteristics were used: **job-related illness behavior** [25, 29, 30] and **perfectionism** [25, 29, 30]. These two constructs were measured with two subscales from the Work Reintegration Questionnaire (WRQ), which is a Dutch validated questionnaire [29, 30]. Both subscales consist of multiple statements which are answered on a 4-point Likert scale (1 = disagree, 2 = somewhat agree, 3 = quite agree, 4 = completely agree). The WRQ scales were dichotomized based on norm scores [29]. The illness behavior scale ranges from 10 to 40 and was dichotomized, with scores above 34 referring to high illness behavior. The perfectionism scale ranges from 12 to 48 and was dichotomized, with scores above 39 referring to high perfectionism.

Disorder-related characteristics

The following disorder-related characteristics were used: **duration of complaints** [11, 31], **pain intensity** [20, 22, 23, 32], **widespread pain** [21, 22, 33], **level of disability** [20, 22, 34, 35], and **perceived health** [22, 23]. Duration of complaints was dichotomized into "subacute" (duration of complaints 3 to 6 months) and "chronic" (more than six months) complaints [31]. Pain intensity was assessed on a 11-point Likert scale, as the mean pain score in the preceding week, where 0 denoted no pain and 10 denoted worst possible pain. Pain intensity was dichotomized into "high pain score" (score of ≥ 7) versus "medium/low pain score" (score of ≤ 6) [2]. Widespread pain was dichotomized into "yes" or "no." Widespread pain was defined as "yes," if pain in the upper extremities (arm, hand, or wrist), lower extremities (hip, knee, ankle, or foot) and axial skeletal pain (back) was present [36].

Level of disability was measured with the Pain Disability Index (PDI) [37], which is a 7-item questionnaire that measures self-reported pain-related disability. The

PDI measures seven dimensions: family/home responsibilities, recreation, social activity, occupation, sexual behavior, self-care, and life support activity on a 0-10 scale (0 denotes “no disability” and 10 denotes “maximum disability”). Total scores range from 0-70, with higher scores reflecting higher level of disability. The level of disability score was dichotomized based on the median. Perceived health was assessed with a single health status item obtained from the RAND-36 [38, 39]: “What do you think about your health in general?,” with five answer categories, ranging from “excellent” to “bad.” Perceived health was dichotomized into good health (“excellent,” “very good,” and “good”) and moderate health (“moderate,” “bad”).

Work-related characteristics

The following work-related characteristics were used: **RTW expectation** [22-25, 32, 40-42], **sick leave duration** [21, 22, 43, 44], **working status** [20, 22, 35, 45], **job strain** [27], and **job dissatisfaction** [24, 46]. RTW expectation was assessed on a 0-10 scale, with patients rating the certainty that they will be working in six months, where 0 represents “Not at all certain” to 10 “Extremely certain.” We dichotomized this item into negative RTW expectancy (score 0-5) and positive RTW expectancy (score 6-10). Sick leave duration was assessed with the sick leave long item of the iPCQ-VR questionnaire [17]. We dichotomized this item into long-term sick leave or not (“yes” = absenteeism for six weeks or more; “no” = absenteeism for less than six weeks). The decision to consider a period of six weeks’ sick leave in this study was based on Dutch social security legislation [47]. Working status was assessed with the working status item of the iPCQ-VR [17]. We dichotomized this item into “full sick leave” and “part-time sick leave.” Job strain and job dissatisfaction were measured with two subscales of the WRQ, which were dichotomized based on norm scores [31]. The job strain scale ranges from 7 to 28 and was dichotomized, with scores above 17 referring to high job strain. The job dissatisfaction scale ranges from 12 to 48 and was dichotomized, with scores above 30 referring to high job dissatisfaction.

Statistical analyses

All analyses were performed using SPSS Statistics for Windows, version 23.0 (2015), IBM Corp., Armonk, NY. The analyses were performed in four steps. In the first step, univariate logistic regression analyses were performed for all independent variables, with work participation as the dependent variable. In the second step, multivariate logistic regression was performed. We applied a

forward selection procedure, with type of intervention as the fixed independent (starting) variable in the model and the independent variables with a p-value of ≤ 0.10 obtained from the univariate analyses (Step 1). Work participation was the dependent variable. We used a p-value of 0.10 for the forward procedure.

In step three, we examined whether confounding variables were present in the first round of the multivariate regression analyses. If the regression coefficient of the *type of intervention* variable increased or decreased ≥ 10%, we considered the independent variable as a confounder. Based on the available evidence, we assumed a priori that RTW expectation [22-25, 32, 40-42], work status [20, 22, 35, 45], and sick leave duration [21, 22, 43, 44] were potential confounders. In the fourth and final step, interaction effects between possible confounders and the dependent variable of work participation were examined using a p-value of < 0.05. Of the final models, model fit was performed based on Hoseman and Lemershow [48]. We report odds ratios, 95% confidence intervals of odds ratios, and p-values. Insight about the relationship between type of intervention and the dependent variable (i.e., work participation) was provided by calculating the proportion of achieved/not achieved work participation and descriptive statistics, separated for type of intervention. We performed the main analyses with complete cases at T0, T1, and T2.

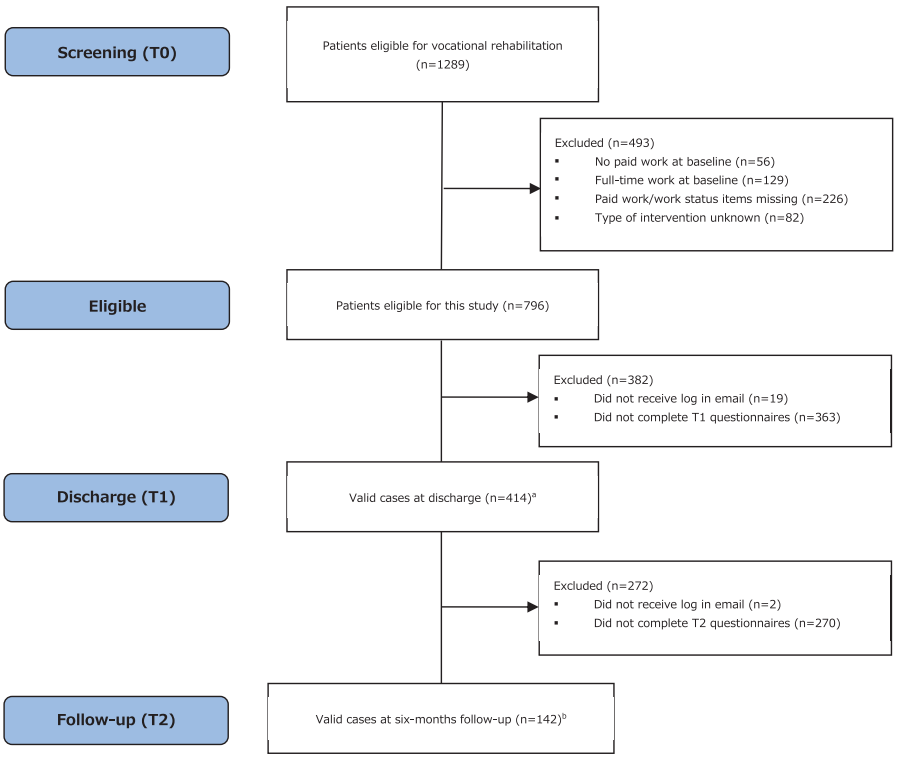
Missing data

Based on earlier (interim) analyses, it was expected that a high proportion of missing data due to loss to follow-up would be present in the dataset, especially for the complete cases. The missing data mechanism (i.e., missing complete at random [MCAR] or missing at random [MAR] [49]) was analyzed by conducting a T-test and Little MCAR tests. We also conducted two additional analyses to explore the influence of missing data on the statistical models. The first additional analyses concerned valid cases on discharge. These patients only completed questionnaires at baseline and discharge. The second additional analyses concerned valid cases on six-months follow-up. These patients only completed questionnaires at baseline and six-months follow-up. For these additional analyses, we followed the same procedure as we had done with the complete cases. A priori, we expected no difference between the final models, confounders, or interaction effects between the complete cases and the additional analyses; however, we did expect smaller confidence intervals and, consequently, a greater likelihood that they would reach statistical significance.

Results

Out of 796 eligible patients, a total of 142 (18%) completed questionnaires at all time points. Of these, 37 (26%) received VR and 105 (74%) VR+. Figure 1 shows a flowchart of the participant inclusion and reasons for dropout. The missing data mechanism for T1 and T2 was missing at random. The sample characteristics of both programs are presented in Table 1.

Figure 1. Flow chart of participants in this study



[a] N=414 patients (52%) completed the discharge questionnaires, but not the six-month follow-up questionnaires. Additional analyses were performed on this subgroup.
[b] N=200 patients (25%) completed the six-month follow-up questionnaires, but not the discharge questionnaires. Additional analyses were performed on this subgroup.

Table 1. Baseline characteristics of the study population (complete cases)

	Complete cases (N=142)	
	VR (N=37)	VR+ (N=105)
	Mean (SD) or %	Mean (SD) or %
Age (years), mean	46.7 (11.8)	47.2 (11.4)
≥ 51 years (%)	53	46
Gender (% female)	54	65
Education[a]		
Low	30	21
Medium	43	41
High	24	30
Other	3	9
Contract (hours/week)	30.9 (11.0)	30.1 (8.8)
Work status		
Part-time sick leave	51	51
Full sick leave	49	49
Sick leave > 6 weeks (% yes)	46	50
Widespread pain (% yes)	24	15
Duration of complaints		
< 6 months	24	21
0.5-1 year	35	26
1-2 years	16	22
2-5 years	3	20
More than 5 years	22	11
Perceived health (% good)	61	59
Pain intensity (0-10)[b]	5.6 (2.4)	5.2 (2.2)
≥ score 7	46	39
Level of disability (PDI 0-70)[c]	37.7 (10.8)	33.8 (12.3)
≥ score 37[d]	49	47
RTW expectancy (0-10)[e]	5.4 (3.1)	6.8 (2.5)
Median	5	7
≥ score 6	47	68
Job strain (7-28)	14.2 (5.1)	15.8 (5.4)
≥ score 18	30	33
Job dissatisfaction (12-48)	24.0 (8.8)	22.3 (7.3)
≥ score 31	19	13
Perfectionism (12-48)	35.7 (7.1)	36.1 (6.3)
≥ score 40	11	5
Job-related illness behavior (10-40)	32.8 (5.2)	31.5 (6.2)
≥ score 35	49	39

SD standard deviation; PDI, pain disability index; RTW, return to work
[a] Education category 'other' not taken into account. Therefore, total percentage may deviate from 100%
[b] 0=no pain, 10=worst possible pain
[c] 0=no disability, 70=maximum disability
[d] Median of total sample was 36
[e] 0=not at all certain, 10=extremely certain

Work participation

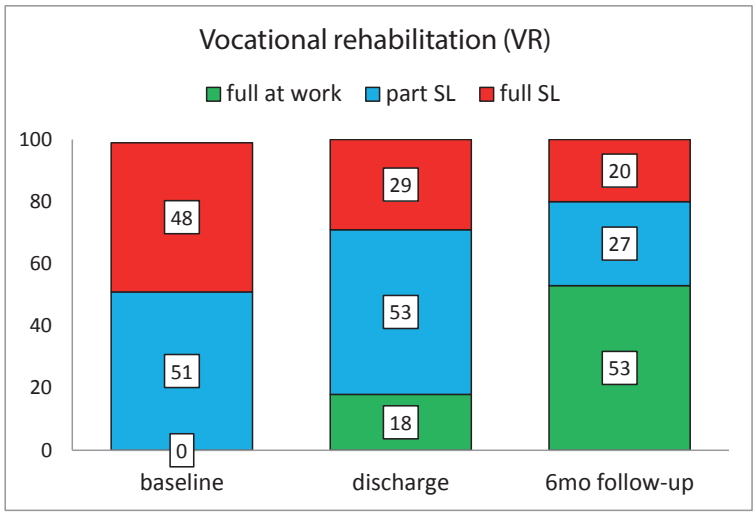
At discharge from vocational rehabilitation, 50% of participants in the VR program and 55% in the VR+ program achieved work participation. At six-months follow-up, 56% of participants in the VR program and 69% in the VR+ program had achieved work participation. The mean number of hours working per week and the working status proportions at each time point for both programs are presented in Table 2 and Figure 2. A non-parametric Mann Whitney U-test showed non-significant differences in working hours per week between VR and VR+ at each time point.

Table 2. Working hours per week for both intervention programs and for the subgroups that achieved/did not achieve work participation at baseline, discharge, and six-months follow-up

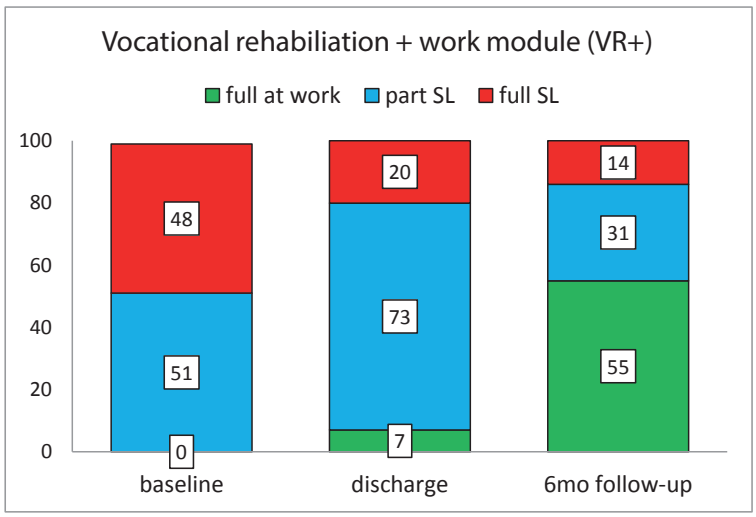
	VR	VR+	Work participation achieved§	Work participation not achieved¥
	Working hours: mean (SD)			
Screening (T0)	6.7 (8.7)	8.0 (9.3)	5.8 (8.3)	9.8 (9.7)
Discharge (T1)	14.2 (13.2)	12.7 (10.4)	18.3 (10.2)	6.6 (8.2)
Difference T1-T0	6.7 (12.5)*	4.6 (9.9)*	12.5 (8.8)*	-3.3 (4.4)*
Follow-up 6 months (T2)	18.0 (15.4)	19.8 (14.0)	27.5 (9.2)	3.0 (6.3)
Difference T2-T0	10.6 (18.3)*	11.7 (14.7)*	20.2 (10.7)*	-5.4 (8.1)*

VR, vocational rehabilitation; VR+, vocational rehabilitation + work module; SD, standard deviation
§ participants who worked at least one hour or more per week at T1/T2 compared to T0
¥ participants who worked the same working hours per week or less at T1/T2 compared to T0
* Significant ($p < 0.05$)

Figure 2. Working status proportions at baseline, discharge, and six-months follow-up



SL, sick leave



SL, sick leave

Regression analyses

The results of the univariate logistic regression analysis are presented in Table 3. The type of intervention had a non-significant relationship to the achievement of work participation at discharge (OR 1.2, p = 0.62) and six-months follow-up (OR 1.8, p = 0.14). The analyses of confounding variables in the relationship

between type of intervention and work participation are presented in Appendix 1. The results of the final multivariate logistic models are presented in Table 4. The type of intervention was not significantly associated with work participation at discharge (OR 1.0, p = 0.99) or six-months follow-up (OR 1.3, p = 0.52). RTW expectation was the only independent factor at discharge (OR 2.5, p = 0.02) and follow-up (OR 2.8, p = 0.01), and a significant confounder at both time points (Appendix 1). No significant interactions were found (results available upon request).

Table 3. Relationship of independent variables with work participation, univariate unadjusted analyses at discharge and six-months follow-up

	Reference category	Discharge		Six-months follow-up	
		Complete cases (N=142)		Complete cases (N=142)	
		P-value	OR (CI 95%)	P-value	OR (CI 95%)
Type of intervention	VR	0.62	1.2 (0.6-2.6)	0.14	1.8 (0.8-3.9)
Pain intensity	Score 7-10	0.85	1.1 (0.5-2.1)	**0.05**	**2.0 (1.0-4.1)[a]**
Widespread pain	Yes	0.48	0.7 (0.3-1.8)	**0.04**	**2.5 (1.0-6.0)[b]**
Perceived health	Good	0.26	1.5 (0.7-2.9)	0.53	0.8 (0.4-1.6)
Age	51-65 years	0.65	0.8 (0.4-1.7)	0.92	1.0 (0.5-2.0)
Gender	Female	0.21	1.6 (0.8-3.1)	0.95	1.0 (0.5-2.1)
Job-related illness behavior	Score 35-40	0.84	0.9 (0.5-1.8)	0.19	1.6 (0.8-3.2)
Perfectionism	Score 40-48	0.89	0.9 (0.2-3.5)	0.52	1.6 (0.4-6.1)
Job strain	Score 18-28	**0.10**	**0.5 (0.3-1.1)**	0.49	1.3 (0.6-2.7)
Job dissatisfaction	Score 31-48	**0.06**	**0.4 (0.1-1.1)**	0.19	0.5 (0.1-1.5)
Sick leave duration	>6 weeks	0.81	1.1 (0.6-2.1)	0.70	1.1 (0.6-2.3)
Duration of complaints	≤6 months	0.87	0.9 (0.4-2.1)	0.49	0.7 (0.3-1.8)
RTW expectation	Score 0-5	**0.03**	**2.1 (1.1-4.3)**	**0.00**	**3.1 (1.5-6.5)**
Level of disability	Score 37-70	0.34	1.4 (0.7-2.7)	**0.09**	**0.8 (0.9-3.7)**
Education, low	NA	0.11	NA	0.37	NA
Education, medium	Low	0.16	0.5 (0.2-1.3)	0.34	0.6 (0.3-1.6)
Education, high	Low	0.67	1.2 (0.5-3.1)	0.79	1.1 (0.4-3.2)
Working status	Full sick leave	**0.04**	**0.5 (0.3-1.0)[c]**	0.19	0.6 (0.3-1.3)

P-value of ≤ 0.10 in bold
[a] original value lower bound: 1.03
[b] original value lower bound: 1.00
[c] original value upper bound: 0.97

Additional analyses

Baseline characteristics of the additional analyses on discharge (n=414) and at six-months follow-up (n=200) are presented in Appendix 2. There were no substantial differences between the baseline characteristics of the complete cases and the additional analyses. Regarding the descriptive statistics of the primary outcome, the additional analyses showed the same pattern as the complete cases. Regarding the univariate analyses, the additional analyses revealed different significant variables (p-value ≤ 0.10) from the complete cases (Appendix 3). The final multivariate regression model of the additional analyses at discharge included working status as a borderline significant factor (p = 0.04, and value 1 not in 95% CI) related to work participation (Appendix 4). In contrast, in the complete cases set, working status was borderline non-significant (p = 0.05, and value 1 in 95% CI) at this time point. The final multivariate regression model of the additional analyses at six-months follow-up included widespread pain as a significant factor related to work participation (Appendix 4).

Table 4. Multivariate analyses with type of intervention (VR+, VR) as fixed variable

	Discharge (N=142)		
	Reference category	P-value	OR (CI 95%)
Type of intervention	VR	0.99	1.0 (0.4-2.3)
RTW expectation	**Score 0-5**	**0.02**	**2.5 (1.2-5.3)**
Working status	Full sick leave	0.05[a]	0.5 (0.2-1)[b]
Job dissatisfaction	Score 31-48	0.07	0.4 (0.1-1.1)
Job strain	Score 18-28	0.24	0.6 (0.3-1.4)

P-value of ≤ 0.05 in bold
[a] Original value: 0.050
[b] Original value lower bound: 1.001

	Six-months follow-up (N=142)		
	Reference category	P-value	OR (CI 95%)
Type of intervention	VR	0.52	1.3 (0.6-3.1)
RTW expectation	**Score 0-5**	**0.01**	**2.8 (1.3-5.9)**
Widespread pain	Yes	0.11	2.2 (0.9-5.5)
Level of disability	Score 37-70	0.34	1.4 (0.7-3.1)

P-value of ≤ 0.05 in bold

Discussion

We hypothesized that patients who received VR+ would have greater odds of achieving work participation compared to patients who received VR. Our hypothesis was not proven. At first sight, the main finding of this study does not appear to be consistent with the strong recommendations of various systematic reviews to include work components to optimize work participation [4, 5, 7, 8, 12, 50, 51].

However, other studies compared multi-domain programs with single-component programs or care as usual [4, 5, 7-9], which complicates comparison of the findings of the present study with them because we compared two multi-component programs. A retrospective cohort study conducted in Canada showed that patients who completed a multimodal pain program that included RTW coordination had 3.4 higher odds of returning to work compared with patients who received the multimodal program without RTW coordination [11]. However, this study did not correct for RTW expectancy.

Based on the present study, and many others [22-25, 32, 40-42], it is clear that RTW expectation is an important confounder in the relationship between an intervention program and a focus on improving work participation. Another RCT study conducted in Norway in patients with neck and back pain showed similar results to our study, namely no significant difference between a group who took part in a multidisciplinary program that included a work focus and a control group who only took part in a multidisciplinary program [10]. One disadvantage of that study, however, was that for the multidisciplinary work-focused group it was not possible to intervene at the workplace due to regulations in Norway. Thus, these results are not directly comparable with those of our study.

In the present study, the proportion of patients at work (full-time or part-time) at six-months follow-up was VR 80% and VR+ 86%. These proportions are slightly higher compared to multi-domain VR described by others, who showed mean work participation proportions of 65% ± 11% [52-58]. In addition, in the present study, the proportion of patients at work full-time at six-months follow-up was VR 53% and VR+ 55%, which is similar to the full-time work proportions reported in other multi-domain VR studies, namely 52% ± 16% [59-63]. In summary, the impact on full-time work participation of the present study, which was performed within clinical practice, was similar to other studies in different countries which were performed in a controlled setting.

Within the Dutch social security system, the employer has a mandatory role in offering modified work. All patients in this study had been offered this in some form, including those in the VR group. In practice, therefore, the contrast between VR and VR+ was smaller than suggested, which may provide an additional plausible explanation for the lack of difference between the groups. The results may thus also provide confirmation, rather than mere falsification of the hypothesis, that work modifications are in fact a core element of VR [4]. How the three core elements (Box 1) should be delivered optimally, however, may depend on country-specific system characteristics and further study.

Strengths and limitations

One strength of a retrospective study is its observational character, as the researcher is able to observe what actually happens or naturally occurs in practice. This is a great advantage in terms of adaptation for professionals. In addition, in our case, it was possible to correct for many independent (potentially confounding) variables which were clustered a priori based on the biopsychosocial model. This increases knowledge of which factors are important to take into account in research and clinical practice. Based on additional analyses, it was possible to detect the influence of more power on the logistic models. This increased the robustness of our findings.

One limitation of a retrospective cohort design is that the intended intervention is less controllable, which may bias the results. In our case, contamination bias between the two programs could have occurred. Patients from both intervention groups were undertaking rehabilitation together. Patients who only participated in the VR program probably obtained information from patients who completed the VR+ program and from the RTW coordinators during group meetings or coffee breaks. Because 3 out of 4 patients received the VR+ program, the chance of contamination bias, resulting in a lack of contrast, was high.

Selection bias may also have occurred, as the type of program a patient participated in was dependent on the employer's willingness to pay for the additional work module. However, at baseline there were no substantial differences between job dissatisfaction and job strain between the VR+ and VR groups. There were probably other factors which influenced the outcomes of the additional work module. From the beginning, it appeared that the VR+ group would have higher odds of achieving work participation compared to the VR group, due to differences in a number of variables: the VR group was

less educated, had a higher proportion of widespread pain, higher pain scores, higher disability scores, and lower RTW expectancy. However, almost all of the independent variables selected a priori were not included or did not contribute to the final multivariate models. The only significant independent variable (and also confounder) in the final multivariate models at discharge and six-months follow-up was RTW expectation. Because selection bias on RTW expectation did not result in a positive association of VR+ and work participation, we assume that the baseline differences between both VR groups did not introduce bias into the results of this study. One final limitation was a high proportion of loss to follow-up, which negatively influenced the sample size of the complete cases (n=142). However, because the results of the additional analyses with larger samples were similar, we assume our findings were not influenced by low power.

Methodological considerations

One methodological consideration with respect to our study concerns the operationalization of the dependent variable of work participation. To detect the influence of our cut-off choice on the reported results, we repeated the univariate and multivariate (if necessary) analyses of the three datasets used in this study. For these additional analyses (not reported; available upon request), we used values ranging from ≥ 2 working hours to ≥ 20 working hours as the cut off for the achievement of work participation. The results showed the same non-significant relationship between type of intervention and the achievement of work participation. This was also observed when the achievement of work participation was operationally defined as full return to work (yes/no). We conclude that our findings would not differ substantially if full-time at work was the dependent variable.

Clinical implications

This study found no significant difference between the effects of VR with or without the addition of a work module on work participation at discharge and six-months follow-up. Both programs showed beneficial RTW rates at six-months follow-up, which is an important message for clinical practice. There was a non-significant, but probably clinically relevant, difference on full sick leave rates at six-months follow-up between both groups (VR+ 14%, VR 20%). Patients, professionals, managers, employers, and policymakers should consider whether this difference suggests that it is worthwhile to add a work module to VR. Before a patient starts VR, it might be advisable to discuss with them which work

components have already been performed at their company, or which steps might be expected during the intervention period, and use this information to decide with them whether a work module should be added to VR. Another implication for practitioners is to take RTW expectations into account before the start of an interdisciplinary VR program, since our study showed that patients with positive RTW expectations had three times higher odds of responding successfully after VR (independent of type of program).

Future directions

In line with the previous point, we recommend that future research should always assess RTW expectations at baseline and correct for this variable during the analyses. Another future direction for research would be to execute return on investment analyses on the added value of work modules when nested in VR. This information is important for those who are asked to reimburse these modules.

Conclusion

This study found no significant difference between interdisciplinary VR programs implemented with or without an additional work module. Both programs were beneficial in improving work participation of sick-listed employees with CMP. Return to work expectations had a strong and significant relationship to the achievement of work participation.

Compliance with Ethical Standards

Funding

No commercial sponsorship was involved in designing or conducting the study.

Conflict of interest

Author TB, author JvV, author CvB, author MFD, and author MR declare that they have no conflict of interest.

Ethical approval

All procedures performed were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The Medical Ethical Committee of the Academic Medical Center, Amsterdam, the Netherlands, authorized this study and decided that a full application was not required.

Informed consent

Informed consent was obtained from all individual participants included in the study.

References

1. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain. 2006 May;10(4):287-333.
2. Koke AJ, Smeets RJ, Schreurs KM, van Baalen B, de Haan P, Remerie SC, et al. Dutch Dataset Pain Rehabilitation in daily practice: Content, patient characteristics and reference data. Eur J Pain. 2017 Mar;21(3):434-44.
3. Sickness, Disability and Work: Breaking the Barriers. A synthesis of findings across OECD countries. OECD 2010.
4. Cullen KL, Irvin E, Collie A, Clay F, Gensby U, Jennings PA, et al. Effectiveness of Workplace Interventions in Return-to-Work for Musculoskeletal, Pain-Related and Mental Health Conditions: An Update of the Evidence and Messages for Practitioners. J Occup Rehabil. 2018 Mar;28(1):1-15.
5. Palmer KT, Harris EC, Linaker C, Barker M, Lawrence W, Cooper C, et al. Effectiveness of community- and workplace-based interventions to manage musculoskeletal-related sickness absence and job loss: a systematic review. Rheumatology (Oxford). 2012 Feb;51(2):230-42.
6. Waddell G, Burton A, Kendall N. Vocational rehabilitation: what works, for whom, and when? The Stationery Office 2013.
7. Waterschoot FP, Dijkstra PU, Geertzen JH, Reneman MF. Dose or content? Effectiveness of pain rehabilitation programs for patients with chronic low back pain: A systematic review. Pain. 2014 Jan;155(1):179-89.
8. van Vilsteren M, van Oostrom SH, de Vet HC, Franche RL, Boot CR, Anema JR. Workplace interventions to prevent work disability in workers on sick leave. Cochrane Database Syst Rev. 2015 Oct 5(10):CD006955.
9. Kamper SJ, Apeldoorn AT, Chiarotto A, Smeets RJ, Ostelo RW, Guzman J, et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain. Cochrane Database Syst Rev. 2014 Sep 2(9):CD000963.
10. Myhre K, Marchand GH, Leivseth G, Keller A, Bautz-Holter E, Sandvik L, et al. The effect of work-focused rehabilitation among patients with neck and back pain: a randomized controlled trial. Spine (Phila Pa 1976). 2014 Nov 15;39(24):1999-2006.
11. Hamer H, Gandhi R, Wong S, Mahomed NN. Predicting return to work following treatment of chronic pain disorder. Occup Med (Lond). 2013 Jun;63(4):253-9.
12. Schaafsma F, Schonstein E, Whelan KM, Ulvestad E, Kenny DT, Verbeek JH. Physical conditioning programs for improving work outcomes in workers with back pain. Cochrane Database Syst Rev. 2010 Jan 20(1):CD001822.
13. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008 Apr;61(4):344-9.
14. Beemster T, van Bennekom C, van Velzen J, Reneman M, Frings-Dresen M. The interpretation of change score of the pain disability index after vocational rehabilitation is baseline dependent. Health Qual Life Outcomes. 2018 Sep 14;16(1):182.

15. Wet verbetering poortwachter (The Law on Gate keeping Disability Insurance). Accessed 22 January 2018; Available from: https://www.arboportaal.nl/onderwerpen/wet-verbetering-poortwachter.
16. Beemster TT, van Velzen JM, van Bennekom CA, Frings-Dresen MH, Reneman MF. Cost-effectiveness of 40-hour versus 100-hour vocational rehabilitation on work participation for workers on sick leave due to subacute or chronic musculoskeletal pain: study protocol for a randomized controlled trial. Trials. 2015 Jul 28;16:317.
17. Beemster TT, van Velzen JM, van Bennekom CAM, Reneman MF, Frings-Dresen MHW. Test-Retest Reliability, Agreement and Responsiveness of Productivity Loss (iPCQ-VR) and Healthcare Utilization (TiCP-VR) Questionnaires for Sick Workers with Chronic Musculoskeletal Pain. J Occup Rehabil. 2018 Mar 16.
18. Waddel G, Burton AK. Concepts of rehabilitation for the management of common health problems. The Stationery Office 2004.
19. Reneman MF, Beemster TT, Edelaar MJ, van Velzen JM, van Bennekom C, Escorpizo R. Towards an ICF- and IMMPACT-based pain vocational rehabilitation core set in the Netherlands. J Occup Rehabil. 2013 Dec;23(4):576-84.
20. Boonstra AM, Reneman MF, Waaksma BR, Schiphorst Preuper HR, Stewart RE. Predictors of multidisciplinary treatment outcome in patients with chronic musculoskeletal pain. Disabil Rehabil. 2015;37(14):1242-50.
21. Oyeflaten I, Lie SA, Ihlebaek CM, Eriksen HR. Prognostic factors for return to work, sickness benefits, and transitions between these states: a 4-year follow-up after work-related rehabilitation. J Occup Rehabil. 2014 Jun;24(2):199-212.
22. Selander J, Marnetoft SU, Bergroth A, Ekholm J. Return to work following vocational rehabilitation for neck, back and shoulder problems: risk factors reviewed. Disabil Rehabil. 2002 Sep 20;24(14):704-12.
23. Steenstra IA, Munhall C, Irvin E, Oranye N, Passmore S, Van Eerd D, et al. Systematic Review of Prognostic Factors for Return to Work in Workers with Sub Acute and Chronic Low Back Pain. J Occup Rehabil. 2017 Sep;27(3):369-81.
24. Opsahl J, Eriksen HR, Tveito TH. Do expectancies of return to work and Job satisfaction predict actual return to work in workers with long lasting LBP? BMC Musculoskelet Disord. 2016 Nov 17;17(1):481.
25. Volker D, Zijlstra-Vlasveld MC, Brouwers EP, van Lomwel AG, van der Feltz-Cornelis CM. Return-to-Work Self-Efficacy and Actual Return to Work Among Long-Term Sick-Listed Employees. J Occup Rehabil. 2015 Jun;25(2):423-31.
26. Fliesser M, De Witt Huberts J, Wippert PM. The choice that matters: the relative influence of socioeconomic status indicators on chronic back pain- a longitudinal study. BMC Health Serv Res. 2017 Dec 2;17(1):800.
27. Haveraaen LA, Skarpaas LS, Berg JE, Aas RW. Do psychological job demands, decision control and social support predictreturn to work three months after a return-to-work (RTW) programme? The rapid-RTW cohort study. Work. 2015;53(1):61-71.
28. Kaikkonen R, Harkanen T, Rahkonen O, Gould R, Koskinen S. Explaining educational differences in sickness absence: a population-based follow-up study. Scand J Work Environ Health. 2015 Jul;41(4):338-46.
29. Vendrig A. De vragenlijst arbeidsreïntegratie (The Work Reintegration Questionnaire). Diagnostiek-Wijzer. 2005;8:27-39.

30. Vendrig L, van Hove M, van Meijel M, Donceel P. Voorspellen van de verwachte verzuimduur met de Vragenlijst ArbeidsReïntegratie (VAR) (Predicting the expected duration of absence with the Work Reintegration Questionnaire (WRQ). TBV. 2011;19(1):7-13.
31. Asih S, Neblett R, Mayer TG, Gatchel RJ. Does the Length of Disability between Injury and Functional Restoration Program Entry Affect Treatment Outcomes for Patients with Chronic Disabling Occupational Musculoskeletal Disorders? J Occup Rehabil. 2018 Mar;28(1):57-67.
32. Besen E, Young AE, Shaw WS. Returning to work following low back pain: towards a model of individual psychosocial factors. J Occup Rehabil. 2015 Mar;25(1):25-37.
33. van der Zee-Neuen A, Putrik P, Ramiro S, Keszei A, de Bie R, Chorus A, et al. Work outcome in persons with musculoskeletal diseases: comparison with other chronic diseases & the role of musculoskeletal diseases in multimorbidity. BMC Musculoskelet Disord. 2017 Jan 10;18(1):10.
34. Storheim K, Brox JI, Holm I, Bo K. Predictors of return to work in patients sick listed for sub-acute low back pain: a 12-month follow-up study. J Rehabil Med. 2005 Nov;37(6):365-71.
35. van Hooff ML, Spruit M, O'Dowd JK, van Lankveld W, Fairbank JC, van Limbeek J. Predictive factors for successful clinical outcome 1 year after an intensive combined physical and psychological programme for chronic low back pain. Eur Spine J. 2014 Jan;23(1):102-12.
36. Leveille SG, Bean J, Ngo L, McMullen W, Guralnik JM. The pathway from musculoskeletal pain to mobility difficulty in older disabled women. Pain. 2007 Mar;128(1-2):69-77.
37. Pollard CA. Preliminary validity study of the pain disability index. Percept Mot Skills. 1984 Dec;59(3):974.
38. Hays RD, Morales LS. The RAND-36 measure of health-related quality of life. Ann Med. 2001 Jul;33(5):350-7.
39. Zee Kvd, Sanderman R. Het meten van de algemene gezondheidstoestand met de RAND-36, een handleiding (Measuring the general health status with the RAND-36, a manual). 2nd ed. Groningen, The Netherlands: RuG, research institute SHARE; 2012.
40. Black O, Keegel T, Sim MR, Collie A, Smith P. The Effect of Self-Efficacy on Return-to-Work Outcomes for Workers with Psychological or Upper-Body Musculoskeletal Injuries: A Review of the Literature. J Occup Rehabil. 2018 Mar;28(1):16-27.
41. Ebrahim S, Malachowski C, Kamal El Din M, Mulla SM, Montoya L, Bance S, et al. Measures of patients' expectations about recovery: a systematic review. J Occup Rehabil. 2015 Mar;25(1):240-55.
42. Hallegraeff JM, Krijnen WP, van der Schans CP, de Greef MH. Expectations about recovery from acute non-specific low back pain predict absence from usual work due to chronic low back pain: a systematic review. J Physiother. 2012;58(3):165-72.
43. Reis RJ, Utzet M, La Rocca PF, Nedel FB, Martin M, Navarro A. Previous sick leaves as predictor of subsequent ones. Int Arch Occup Environ Health. 2011 Jun;84(5):491-9.
44. Roelen CA, Koopmans PC, Schreuder JA, Anema JR, van der Beek AJ. The history of registered sickness absence predicts future sickness absence. Occup Med (Lond). 2011 Mar;61(2):96-101.

45. Viikari-Juntura E, Virta LJ, Kausto J, Autti-Ramo I, Martimo KP, Laaksonen M, et al. Legislative change enabling use of early part-time sick leave enhanced return to work and work participation in Finland. Scand J Work Environ Health. 2017 Sep 1;43(5):447-56.

46. Stapelfeldt CM, Christiansen DH, Jensen OK, Nielsen CV, Petersen KD, Jensen C. Subgroup analyses on return to work in sick-listed employees with low back pain in a randomised trial comparing brief and multidisciplinary intervention. BMC Musculoskelet Disord. 2011 May 25;12:112.

47. Dekkers-Sanchez PM, Hoving JL, Sluiter JK, Frings-Dresen MH. Factors associated with long-term sick leave in sick-listed employees: a systematic review. Occup Environ Med. 2008 Mar;65(3):153-7.

48. Hosmer DW, Lemeshow S. Applied logistic regression. New York: John Wiley & Sons; 1989.

49. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009 Jun 29;338:b2393.

50. Hoefsmit N, Houkes I, Nijhuis FJ. Intervention characteristics that facilitate return to work after sickness absence: a systematic literature review. J Occup Rehabil. 2012 Dec;22(4):462-77.

51. Meijer EM, Sluiter JK, Frings-Dresen MH. Evaluation of effective return-to-work treatment programs for sick-listed patients with non-specific musculoskeletal complaints: a systematic review. Int Arch Occup Environ Health. 2005 Aug;78(7):523-32.

52. Bethge M, Herbold D, Trowitzsch L, Jacobi C. Work status and health-related quality of life following multimodal work hardening: a cluster randomised trial. J Back Musculoskelet Rehabil. 2011;24(3):161-72.

53. Brendbekken R, Eriksen HR, Grasdal A, Harris A, Hagen EM, Tangen T. Return to Work in Patients with Chronic Musculoskeletal Pain: Multidisciplinary Intervention Versus Brief Intervention: A Randomized Clinical Trial. J Occup Rehabil. 2017 Mar;27(1):82-91.

54. Harris A, Moe TF, Eriksen HR, Tangen T, Lie SA, Tveito TH, et al. Brief intervention, physical exercise and cognitive behavioural group therapy for patients with chronic low back pain (The CINS trial). Eur J Pain. 2017 Sep;21(8):1397-407.

55. Jousset N, Fanello S, Bontoux L, Dubus V, Billabert C, Vielle B, et al. Effects of functional restoration versus 3 hours per week physical therapy: a randomized controlled study. Spine (Phila Pa 1976). 2004 Mar 1;29(5):487-93.

56. Linton SJ, Boersma K, Traczyk M, Shaw W, Nicholas M. Early Workplace Communication and Problem Solving to Prevent Back Disability: Results of a Randomized Controlled Trial Among High-Risk Workers and Their Supervisors. J Occup Rehabil. 2016 Jun;26(2):150-9.

57. Skouen JS, Grasdal AL, Haldorsen EM, Ursin H. Relative cost-effectiveness of extensive and light multidisciplinary treatment programs versus treatment as usual for patients with chronic low back pain on long-term sick leave: randomized controlled study. Spine (Phila Pa 1976). 2002 May 1;27(9):901-9.

58. Henchoz Y, de Goumoens P, So AK, Paillex R. Functional multidisciplinary rehabilitation versus outpatient physiotherapy for non specific low back pain: randomized controlled trial. Swiss Med Wkly. 2010;140:w13133.

59. Aasdahl L, Pape K, Vasseljen O, Johnsen R, Gismervik S, Halsteinli V, et al. Effect of Inpatient Multicomponent Occupational Rehabilitation Versus Less Comprehensive Outpatient Rehabilitation on Sickness Absence in Persons with Musculoskeletal- or Mental Health Disorders: A Randomized Clinical Trial. J Occup Rehabil. 2017 Apr 11.
60. Bultmann U, Sherson D, Olsen J, Hansen CL, Lund T, Kilsgaard J. Coordinated and tailored work rehabilitation: a randomized controlled trial with economic evaluation undertaken with workers on sick leave due to musculoskeletal disorders. J Occup Rehabil. 2009 Mar;19(1):81-93.
61. Indahl A, Velund L, Reikeraas O. Good prognosis for low back pain when left untampered. A randomized clinical trial. Spine (Phila Pa 1976). 1995 Feb 15;20(4):473-7.
62. Jensen C, Jensen OK, Christiansen DH, Nielsen CV. One-year follow-up in employees sick-listed because of low back pain: randomized clinical trial comparing multidisciplinary and brief intervention. Spine (Phila Pa 1976). 2011 Jul 1;36(15):1180-9.
63. Moll LT, Jensen OK, Schiottz-Christensen B, Stapelfeldt CM, Christiansen DH, Nielsen CV, et al. Return to Work in Employees on Sick Leave due to Neck or Shoulder Pain: A Randomized Clinical Trial Comparing Multidisciplinary and Brief Intervention with One-Year Register-Based Follow-Up. J Occup Rehabil. 2018 Jun;28(2):346-56.

Appendix 1. Confounding analyses of the relationship of type of intervention with work participation at discharge and six-months follow-up (executed for complete cases and additional analyses datasets)

Complete cases: Discharge (N=142)

	Reference category	P-value	OR (CI 95%)	Wald	B unadj.	B adj.	B change (%)
Job strain	**Score 18-28**	**0.10**	**0.5 (0.3-1.1)**	**2.7**	**0.193**	**0.146**	**24**
Job dissatisfaction	**Score 31-48**	**0.06**	**0.4 (0.1-1.0)**	**3.6**	**0.193**	**0.221**	**-15**
RTW expectation	**Score 0-5**	**0.04**	**2.1 (1.05-4.3)**	**4.4**	**0.193**	**0.005**	**97**
Working status	Full sick leave	0.04	0.5 (0.3-1.0)[a]	4.2	0.193	0.206	-7

Confounders in bold
[a] Original value upper limit: 0.97

Complete cases: Six-months follow-up (N=142)

	Reference category	P-value	OR (CI 95%)	Wald	B unadj.	B adj.	B change (%)
Pain intensity	Score 7-10	0.06	2.0 (0.8-3.8)	3.5	0.588	0.542	8
Widespread pain	**Yes**	**0.06**	**2.3 (1.0-5.7)**	**3.5**	**0.588**	**0.518**	**112**
RTW expectation	**Score 0-5**	**0.00**	**3.0 (1.4-6.3)**	**8.4**	**0.588**	**0.341**	**42**
Level of disability	Score 37-70	0.09	1.8 (0.9-3.7)	2.8	0.588	0.594	-1

Confounders in bold

Additional analyses: Discharge (N=414)

	Reference category	P-value	OR (CI 95%)	Wald	B unadj.	B adj.	B change (%)
RTW expectation	**Score 0-5**	**0.00**	**2.7 (1.8-4.1)**	**22.8**	**0.593**	**0.501**	**16**
Education, low	**NA**	**0.26**	**NA**	**2.7**	**0.593**	**0.49**	**17**
Education, medium	Low	0.56	1.2 (0.7-1.9)	0.3	NA	NA	NA
Education, high	Low	0.11	1.6 (0.9-2.7)	2.5	NA	NA	NA
Working status	Full sick leave	0.13	0.7 (0.5-1.1)	2.3	0.593	0.571	4

Confounders in bold

Additional analyses: Six-months follow-up (N=200)

	Reference category	P-value	OR (CI 95%)	Wald	B unadj.	B adj.	B change (%)
Pain intensity	Score 7-10	0.14	1.6 (0.9-2.9)	2.2	0.464	0.431	7
Widespread pain	**Yes**	**0.04**	**2.2 (1.1-4.7)**	**4.4**	**0.464**	**0.412**	**11**
Job-related illness behavior	**Score 35-40**	**0.06**	**1.8 (1.0-3.3)[a]**	**3.6**	**0.464**	**0.377**	**19**
RTW expectation	**Score 0-5**	**0.00**	**3.1 (1.7-6.0)**	**12.8**	**0.464**	**0.231**	**50**
Level of disability	**Score 37-70**	**0.08**	**1.7 (0.9-3.2)**	**3.2**	**0.464**	**0.383**	**18**

Confounders in bold
[a] Original value lower bound: 0.98

Appendix 2. Baseline characteristics of the additional analyses study samples

	Discharge Additional analyses (N=414)		Six-months follow-up Additional analyses (N=200)	
	VR (N=109) Mean (SD) or %	VR+ (N=305) Mean (SD) or %	VR (N=51) Mean (SD) or %	VR+ (N=149) Mean (SD) or %
Age (years), mean	47.2 (11.0)	47.3 (10.5)	46.1 (12.1)	47.7 (10.9)
≥ 51 years (%)	47	45	49	47
Gender (% female)	54	62	51	65
Education[a]				
Low	38	21	26	22
Medium	40	42	46	41
High	19	31	26	30
Other	3	6	2	7
Contract (hours/week)	30.3 (11.5)	30.8 (9.2)	31.9 (10.1)	30.2 (8.9)
Work status				
Part-time sick leave	56	48	49	53
Full sick leave	44	52	51	47
Sick leave > 6 weeks (% yes)	44	57	49	50
Widespread pain (% yes)	21	12	24	17
Duration of complaints				
< 6 months	25	24	28	22
0.5-1 year	50	53	33	25
1-2 years	22	19	14	19
2-5 years	8	14	8	21
More than 5 years	20	13	18	13
Perceived health (% good)	55	55	61	60
Pain intensity (0-10)[b]	5.7 (2.1)	5.2 (2.3)	6.0 (2.2)	5.3 (2.3)
≥ score 7	45	39	54	41
Level of disability (PDI 0-70)[c]	35.5 (11.0)	35.7 (11.9)	39.9 (10.3)	34.0 (12.4)
≥ score 37[d]	43	49	61	46
RTW expectancy (0-10)[e]	5.5 (3.1)	6.5 (2.6)	5.5 (3.0)	6.7 (2.5)
Median	5	7	5	7
≥ score 6	47	66	45	65
Job strain (7-28)	14.6 (5.5)	15.2 (5.2)	14.9 (5.1)	15.9 (5.4)
≥ score 18	31	30	31	35
Job dissatisfaction (12-48)	24.4 (8.0)	22.9 (7.2)	24.6 (8.5)	22.8 (7.7)
≥ score 31	22	15	24	15
Perfectionism (12-48)	34.6 (6.9)	35.2 (6.2)	36.6 (7.0)	35.8 (6.3)
≥ score 40	9	6	8	5
Job-related illness behavior (10-40)	32.5 (5.7)	31.7 (5.8)	33.2 (5.3)	31.6 (5.8)
≥ score 35	43	39	53	40

SD standard deviation; PDI, pain disability index; RTW, return to work
[a] Education category 'other' not taken into account. Therefore, total percentage may deviate from 100%
[b] 0=no pain, 10=worst possible pain
[c] 0=no disability, 70=maximum disability
[d] Median of total sample of the complete cases was 36 (see Table 1)
[e] 0=not at all certain, 10=extremely certain

Appendix 3. Relationship of independent variables with work participation: univariate unadjusted analyses with additional analyses datasets at discharge and six-months follow-up

		Discharge		Six-months follow-up	
		Additional analyses (N=412)		Additional analyses (N=200)	
	Reference category	P-value	OR (CI 95%)	P-value	OR (CI 95%)
Type of intervention	VR	**0.01**	**1.8 (1.2-2.8)**	0.17	1.6 (0.8-3.1)
Pain intensity	Score 7-10	0.11	1.4 (0.9-2.1)	**0.10**	**1.7 (0.9-3.0)**
Widespread pain	Yes	0.56	1.2 (0.7-2.0)	**0.03**	**2.3 (1.1-4.8)**
Perceived health	Good	0.91	1.0 (0,7-1.4)	0.50	0.8 (0.4-1.5)
Age	51-65 years	0.53	1.1 (0.8-1.7)	0.49	0.8 (0.4-1.5)
Gender	Female	0.43	1.2 (0.8-1.7)	0.94	1.0 (0.6-1.9)
Job-related illness behavior	Score 35-40	0.43	1.2 (0.8-1.8)	**0.04**	**1.9 (1.0-3.4)**[a]
Perfectionism	Score 40-48	0.29	0.6 (0.3-1.4)	0.45	1.6 (0.5-5.2)
Job strain	Score 18-28	0.97	1.0 (0.7-1.5)	0.83	1.1 (0.6-2.0)
Job dissatisfaction	Score 31-48	0.67	0.9 (0.5-1.5)	0.31	0.6 (0.3-1.5)
Sick leave duration	>6 weeks	0.95	1.0 (0.7-1.5)	0.30	1.4 (0.8-2.5)
Duration of complaints	≤6 months	0.78	0.9 (0.6-1.5)	0.18	0.6 (0.3-1.3)
RTW expectation	Score 0-5	**0.00**	**2.8 (1.9-4.2)**	**0.00**	**3.3 (1.8-6.1)**
Level of disability	Score 37-70	0.25	1.3 (0.9-1.9)	**0.06**	**1.8 (1.0-3.3)**[b]
Education, low	NA	0.14	NA	0.32	NA
Education, medium	Low	0.39	1.2 (0.8-2.0)	0.34	0.7 (0.3-1.5)
Education, high	Low	**0.05**	**1.7 (1.0-2.9)**	0.73	1.2 (0.5-2.8)
Working status	Full sick leave	**0.09**	**0.7 (0.5-1.1)**	0.28	0.7 (0.4-1.3)

P-value of ≤ 0.10 in bold
[a] original value lower bound: 1.02
[b] original value lower bound: 0.99

Appendix 4. Multivariate analyses with type of intervention as fixed variable (analyses of additional datasets at discharge and six-months follow-up)

	Discharge (N=414)		
	Reference category	P-value	OR (CI 95%)
Type of intervention	VR	0.15	1.4 (0.9-2.3)
RTW expectation	**Score 0-5**	**0.00**	**2.9 (1.9-4.4)**
Education, low	0.29	NA	NA
Education, medium	Low	0.57	1.2 (0.7-1.9)
Education, high	Low	0.13	1.6 (0.9-2.7)
Working status	**Full sick leave**	**0.04**	**0.6 (0.4-1.0)[a]**

P-value of ≤ 0.05 in bold
NA, not applicable
[a] Original value upper bound: 0.97

	Six-months follow-up (N=200)		
	Reference category	P-value	OR (CI 95%)
Type of intervention	VR	0.86	1.1 (0.5-2.2)
RTW expectation	**Score 0-5**	**0.00**	**3.0 (1.5-5.7)**
Widespread pain	**Yes**	**0.02**	**2.7 (1.1-6.3)**
Pain intensity	Score 7-10	0.64	0.8 (0.4-1.8)
Job-related illness behavior	Score 35-40	0.24	1.5 (0.8-3.0)
Level of disability	Score 37-70	0.26	1.5 (0.7-3.0)

P-value of ≤ 0.05 in bold

CHAPTER 8

General Discussion

The overall aim of this thesis is to contribute to quality improvement of vocational rehabilitation (VR) for patients with chronic musculoskeletal pain and reduced work participation. This aim was divided into two parts. Part I aimed to investigate the clinimetric properties of work participation, healthcare usage, and pain-related disability measures. Part II aimed to investigate the relationship between the "dosage" and "content" of VR on work participation. In this chapter, the main findings, interpretation of these findings, and the methodological considerations of Chapters 2-7 are discussed. Recommendations for VR professionals, practice, researchers, and future research are provided. An epilogue ends the chapter.

Main findings

Research question 1: *Which questionnaires should be included in a focused "VR-pain Core Set" that can be used across VR practice in the Netherlands and can examine clinical and cost effectiveness?*

Development of a core set of diagnostic and evaluative measures for patients with CMP and reduced work participation—specifically tailored for use in the context of Dutch VR centers—was described in Chapter 2. The "VR-pain Core Set" consists of items from the following questionnaire tools: EuroQol 5 Dimensions (EQ-5D), Work Ability Index (WAI), productivity and disease questionnaire (PRODISQ, later replaced by the iMTA productivity Cost Questionnaire-Vocational Rehabilitation (iPCQ-VR)), Pain Disability Index (PDI), RAND-36 physical functioning scale, work reintegration questionnaire (WRQ), Numeric Rating Scale (NRS) pain, NRS fatigue, lifting test, Astrand bicycle test or Bruce treadmill test, Trimbos iMTA questionnaire for measuring costs of psychiatric illnesses (TiC-P, later replaced by the TiCP-VR), and the Global Perceived Effect (GPE). Of these, iPCQ-VR, TiCP-VR, and EQ-5D can be used for cost effectiveness purposes.

Research question 2*: What are the clinimetric properties of work participation, healthcare usage, and pain-related disability questionnaires for patients with CMP and reduced work participation in attendance of, and following discharge from, VR in the Netherlands?*

Retest reliability, agreement, and responsiveness of the iPCQ-VR questionnaire, which measures work participation, were assessed in Chapter 3. The iPCQ-VR showed good measurement properties with regard to "working status," "number

of hours working per week," and "long-term sick leave." The measurement properties of "short-term sick leave" and "presenteeism" were poor. The retest reliability and agreement of the TiCP-VR questionnaire, which measures health care usage, were also examined in Chapter 3. "Total health care usage" showed sufficient reliability; however, the "single healthcare usage" items exhibited varying reliability and agreement figures, from very poor to almost perfect reliability and agreement.

Responsiveness and interpretation of change scores for the Pain Disability Index (PDI) questionnaire were examined in Chapter 4. The results showed that the PDI was responsive to real changes in pain-related disability in a sample of patients with CMP and reduced work participation after engaging in VR. Subsamples based on PDI baseline quartile scores also showed adequate responsiveness. Change scores were provided for the total study sample and the subsamples.

Research question 3: *What are the opinions and experiences of patients, professionals, and managers regarding the usefulness and feasibility of "comprehensive" and "less-comprehensive" VR programs?*

The findings of interviews conducted with patients who were included in the RCT (Chapter 5) and who had completed their allocated program are provided in Chapter 6. Patients were allocated to either a 100-hour "comprehensive" VR program (C-VR) or a 40-hour "less-comprehensive" VR program (LC-VR). The main findings were that both programs are considered feasible and generally useful. However, some patients stated that not all of the content was useful, and, in some, content saturation took place. Professionals preferred working with the "C-VR as standard" program, although some disliked its rigid and uniform character. Professionals also felt that the C-VR program was too extensive for some patients and that these patients would likely benefit from the LC-VR program. Several patient factors were identified by professionals that might enhance allocation methods to either C-VR or LC-VR programs. Managers felt that, despite appreciating the relevance of the LC-VR program, implementation of the program would not be financially possible due to the Dutch healthcare system. The overall conclusion from the patients, professionals, and managers was that it is not useful to deliver one VR program for all patients and that treatment should be personalized through the use of quasi-flexible and tailored VR.

Research question 4: *Are patients with CMP and reduced work participation who attended "VR with work module" more likely to achieve work participation than patients who attended "VR without work module?"*

A retrospective cohort study is presented in Chapter 7. This looks at the relationship between VR—with and without an additional work module—on the work participation of patients with CMP and reduced work participation, both at discharge and at six-months follow-up. The results showed that there was no significant difference between a VR program with an additional work module (denoted VR+) and a program without an additional work module (denoted VR), on work participation at both the time points measured. There was a small difference in working status (full and part-time work) at discharge (VR+ 80%; VR 71%) and at six months follow up (VR+ 86%; VR 80%), but these differences were non-significant. Chapter 7 also showed that the variable "return to work expectation" was strongly related to work participation.

Interpretation of findings

The findings presented in Chapters 2-7 are interpreted in this section.

Part I: the clinimetric properties of work participation, healthcare usage, and pain-related disability measures

Work participation

To better contextualize the iPCQ-VR work participation findings presented in Chapter 3, a framework describing five types of work disability—as proposed by Young et al. [1]—can be used. This model consists of *type 0*, at work, no work disability; *type 1*, working, but experiencing health-related work limitations; *type 2*, off work due to health condition; *type 3*, returned to work with work limitations, and *type 4*, withdrawn from the labor force. Various outcome measures can be assessed at each level of work disability (WD) [1]. For instance, productivity, presenteeism, work limitations, and work abilities can be assessed in those with type 1 WD. In type 2, time off work, employee-employer interactions, return to work (RTW) preparations, and work absence recurrence can be measured. For those classified with type 3 WD, time until RTW, time until back at work, time until sustained RTW is achieved, durable RTW, and proportion of time at work (e.g., working status) can all be measured. Labor force participation and vocational status can both be recorded in type 4 WD patients. In sum, there is a

great deal of variation in the possible measurements for work participation and RTW [1, 2]. This variation makes comparison between VR programs difficult [2]. To increase comparability between scientific publications and to enhance clinical usefulness, Young et al. [1] proposed the use of a standardized set of outcome measures that enable trajectory analysis. This should involve multidimensional outcome assessments of a range of variables, taken over extended periods. The iPCQ-VR consists of measurements applicable to all WD types—for instance, presenteeism (types 0, 1, and 3), sick leave days (types 2 and 3), and working status (all WD types)—making it useful for purposes of evaluation and allowing for trajectory analysis [1].

Despite the positive attributes of the iPCQ-VR detailed above, a significant shortcoming of the tool is that it does not include the measurement of *sustainable RTW*, which can be defined as the number of days on wage replacement benefits followed by at least 28 days without receiving these benefits [3]. Sustainable RTW is a frequently used outcome measure in VR research [1, 3, 4], probably because it enhances trajectory analysis. The iPCQ-VR components "sick leave short" and "sick leave long" could, theoretically, be used as proxies by which to assess a sustainable RTW; however, as described in Chapter 3, the measure "sick leave short" showed poor retest reliability, whereas "sick leave long" exhibited only sufficient reliability for use at group level (not on an individual level). Therefore, these two measures cannot be used to provide an adequate assessment of sustainable RTW. Another potential proxy of sustainable RTW could be the "working status" and "number of hours working per week" items of the iPCQ-VR (Chapter 3); however, these measures have the shortcoming that they only afford a recall period of 1 week. To increase recall accuracy, and as a 12-week duration is often used for measurement purposes in clinical studies, research has proposed to extend the recall period of these measures to 2-3 months [5, 6]. This could be achieved by, for example, adding an extra categorical variable to the iPCQ-VR that assesses the length of time that the response on "working status" and "number of hours working per week" items takes (with answer categories from 1 week to 3 months). Future research should study the reliability and validity of such additional variable.

Other shortcomings of the iPCQ-VR are the low retest reliability and responsiveness values for "presenteeism" (Chapter 3), demonstrating that caution must be exercised if applying this concept to clinical practice or research. Higher power studies have demonstrated slightly greater reliability figures [7], but reliability in these studies remains too low for individual-level evaluation purposes. As

such, there is no gold standard for the measurement of presenteeism [6, 8, 9]. A possibility for improving the reliability of presenteeism measures could be to assess presenteeism multiple times in reliability studies. In a study looking at task-specific work ability [10], the reliability of task-specific work ability was found to increase over the use of two (ICC 0.65), three (ICC 0.71), and twelve (ICC 0.86) measurement points. It can be assumed that measurement values for presenteeism exhibit similar variation compared to values for work ability. Therefore, incorporating a greater number of measurement points might more accurately capture variations in presenteeism values, and, as such, provide better reliability figures. Future research should study this hypothesis.

Healthcare utilization

When combined, all TiCP-VR items concerning healthcare utilization showed sufficient reliability and can, therefore, be used at a group level. The single items, however, showed low to moderate reliability, and require further investigation. Several improvements are suggested here for the use of the TiCP-VR in economic studies. First, an increase in the recall period from 1 month to 3 months [5] may be beneficial, since it has been proposed that "collecting data with relatively short recall periods (e.g., a couple of weeks) over a longer period of time may be overly burdensome to participants and may thus increase the risk of missing data and dropout. Therefore, it may be better to maximize completeness at the cost of some recall bias, for example, by using 2- to 3-month recall periods in a trial with a long-term follow-up (≥12 months)" [5: p. 565]. Another improvement might be to measure only generic healthcare usage, not, as is the case with the TiCP-VR, both generic and VR pain specific healthcare usage (Chapter 3). This will increase feasibility since the criterion validity of pain specific healthcare usage items is low [11]. To further increase feasibility, it would also be prudent to delete health care items that are seldom consulted or used by patients. For instance, Chapter 3 shows that the TiCP-VR items "insurance physician," "social worker," "dietician," inpatient "stay in healthcare setting," and "home care" are seldom used. Therefore, these items might be deleted from the TiCP-VR. Given all of these suggested improvements, it may make more sense to instead use the iMTA medical consumption questionnaire (iMCQ) [11, 12] and to adapt this questionnaire to a VR context. The iMCQ was developed in 2013 and measures generic health care usage with a recall period of 3 months [11]. The iMCQ is recommended by the Dutch guideline for health economic evaluations [12]. Future research should look to study the clinimetric properties of a VR adapted version of the iMCQ.

Pain-related disability

Chapter 4 details that, with respect to PDI scores, patients who were less disabled at baseline had to improve by 10% (7 PDI points) in order for the change score to be clinically relevant; in contrast, those patients severely disabled at baseline had to improve by 30% (20 PDI points) in order for the changes to be clinically relevant. Studies on pain, health status, and functional ability score also describe a linear relationship between baseline score and change score at discharge and follow up [13-15]. In order to account for this relationship, previous research has proposed the performance of ''responder analysis" [14, 15]. Responder analysis focuses on the percentage of patients that have reached a relevant change. This has been recommended as a readily interpretable measure that can be of relevance for both researchers and clinicians [14]. Responder analysis was performed in Chapter 4. The results showed that when "improvement" was defined *relatively*, with a change score ≥MIC (minimal important change), the improvement on PDI baseline quartiles ranged between 40-46%. If "improvement" was defined as a PDI change score of ≥1 point, improvements ranged between 65-88%. This example demonstrates that the choice of cutoff point for measuring improvement influences results, and thus the conclusions drawn, regarding changes in pain-related disability after attending VR. This concept should be borne in mind when analyzing or interpreting study results.

Part II: the relationship between the dosage and content of VR on work participation

Relationship dose-content

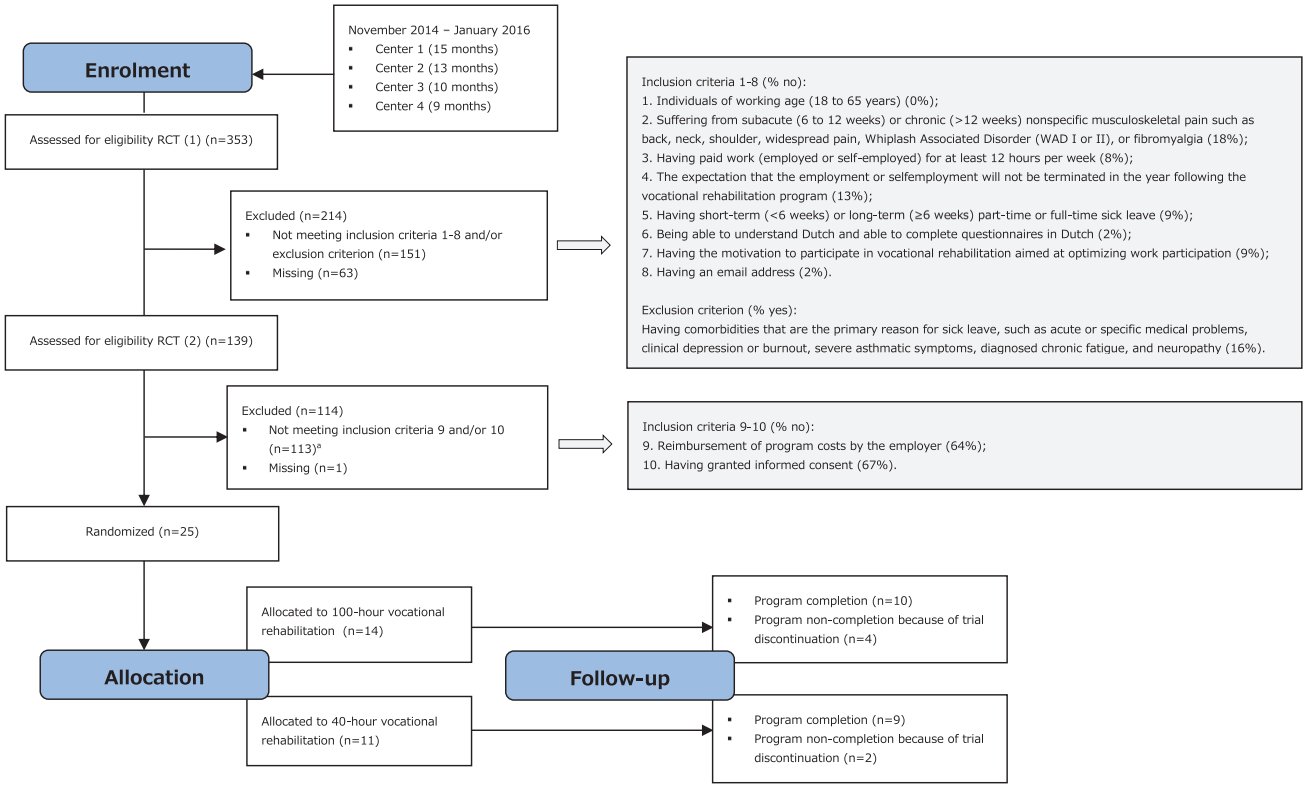
The second aim of this thesis was to investigate the relationship between the *dosage* and *content* of VR on work participation. In order to address this, a multicenter RCT was implemented, in which it was hypothesized that a less-comprehensive 40-hour VR (experimental) program could be non-inferior in regard to work participation, and also cost effective, compared with a comprehensive 100-hour VR (standard care) program (Chapter 5). Due to a low inclusion rate, it was decided to end the RCT early at 1.5 years (Fig. 1 shows a flow chart of the RCT design). Henceforth, it was not possible to draw any conclusions from the RCT concerning the non-inferiority hypothesis.

However, in recent years, a number of similar studies have been published in other countries. All of these studies have shown that simplified, or less-comprehensive, programs are non-inferior in regard to work participation compared with comprehensive programs [16-22]. There are various possible

explanations for these findings. It is very likely that the *type* of dosage model plays a role. In an attempt to improve treatment outcome and efficiency in psychiatric care, researchers compared two dosage models: the "dose-effect model" and the "good-enough-level" (GEL) model [23-25]. The dose-effect model [23] states that a longer duration of therapy leads to better outcomes; however, this will yield diminishing returns, as increasing the number of sessions results in progressively less change. The good-enough-level (GEL) model [23] proposes that either the patient, therapist, or both in conjunction, decide upon a treatment endpoint whenever the treatment outcome is satisfactory; this means that patients who recover faster will have shorter treatments. According to Baldwin et al. [25], "the dose-effect model predicts that rate of change during therapy will not vary as a function of total number of sessions, whereas the GEL model predicts that it will vary (p. 204)." Additional studies in both psychological and psychiatric care have demonstrated that the rate of change over treatment varies as a function of the total number of sessions, which is consistent with the GEL model [23-25]. Baldwin et al. further describe that "rate of change was related to total dose of treatment—small doses were related to relatively fast rates of change, whereas large doses were related to slow rates of change (p. 208)." The dose-effect model, therefore, also resulted in positive changes on relevant outcomes, but at a slower rate compared with the GEL model [23-25]. In sum, there is mixed support in the psychiatric literature for the GEL and dose-effect models [25].

The reported non-inferiority of less-comprehensive programs compared to comprehensive programs, in relation to their impact on work participation, might be partially explained by the similarity of the former to the GEL model; whereas the later could be seen as being similar to the dose-effect model. The insights derived from psychiatric dosage models (presented above) might suppose to use both the GEL and dose-effect model in VR practice to accomplish effective and efficient care. Further—indirect—evidence supports this supposition. As described in Chapter 5, the patients who followed the LC-VR or C-VR programs as part of the RCT, mentioned that content saturation in group education or individual sessions with the psychologists occurred after a couple of weeks, suggesting that a GEL model may be beneficial in this subgroup of patients. A qualitative study in patients with chronic pain who attended pain rehabilitation found comparable results concerning treatment saturation [26]. Another paper [27], in which brief intervention was compared with brief intervention *with* group cognitive behavioral therapy or brief intervention *with* group physical exercise, in patients on sick leave for 2-10 months due to nonspecific low back

Figure 1. Inclusion flow of the multicenter RCT



[a] Inclusion criterion 9: no (n=89), yes (n=50); inclusion criterion 10: no (n=93), yes (n=45)

pain, showed that brief intervention was the superior program in improving work participation. The authors [27] explained this finding as follows: "the lack of significant additional effects could imply that the psychological and physiological elements already had been sufficiently addressed in the brief intervention and that further treatment, therefore, had little impact on the outcome (p. 9)." Finally, a study which compared the use of a multicomponent program with standard care, in patients with chronic back pain, showed higher effectiveness rates on return to work in favor of the multicomponent program [28]. Interestingly, the multicomponent program duration was a maximum of 12 weeks but stopped as soon as return to own or equal work was established, thus reflecting the GEL model.

A significant body of evidence, therefore, suggests the utility of a mixed GEL and dose-effect model in VR practice; however, it should be noted that this evidence comes primarily from monodisciplinary psychiatric care. Moreover, it could be assumed that a mixed GEL and dose-effect model may interfere in the group process, which is often utilized in VR. Moreover, various studies—including Chapter 6—have shown the benefits of rehabilitation in a group and therapeutic discussion with peers; for example, in the provision of social support, understanding of problems, acceptance, developing self-esteem, sharing experiences, and obtaining information from others [29-34]. Potential interference in the group process arises because some patients will leave the group earlier than others; this is a significant factor to account for if a mixed GEL and dose-effect model is to be implemented in VR practice. This is a topic for future research.

Regardless of the type of dosage model used, it is important to consider a number of factors in relation to the aforementioned non-inferiority findings in the literature. First, the research comparing shorter, or less-comprehensive, programs with longer, comprehensive ones [16-22], exhibited a wide range in both the dosage of therapy prescribed (e.g., number of contact hours, frequency, and duration of treatment) and the content of the program. This makes it difficult to develop guidelines and address specific recommendations for stakeholders. A second issue is that these studies have all been conducted in other countries. Research has shown that the generalization of results from a study conducted in a particular country to another country is difficult because of differences in healthcare systems [35-37]. For instance, the majority of comparative studies of relevance to this section were conducted in Scandinavian countries [2]. Since the Scandinavian social welfare model generally provides universally

accessible benefits, application to the Dutch healthcare context is difficult, whereby healthcare insurers and employers have to reimburse the costs of VR and additional work modules. For example, in Chapter 6 managers stated that the implementation of less comprehensive programs was not feasible for Dutch VR centers due to the healthcare system in the Netherlands. The number of non-inferiority studies carried out in Scandinavian countries indicates that their system has greater flexibility in which to test innovative VR programs.

A third consideration is the patient recruitment strategy employed. In a study by Harris et al. [22, 27], patients on sick leave due to chronic low back pain in Norway were recruited on a voluntary basis via a letter sent by the Norwegian Labour and Welfare Administration. The baseline population of that study showed a low mean disability score as compared to various categories of patients with chronic low back pain [27]; this indicates that *less complex* patients were included in the study, probably as a result of the applied recruitment strategy. Other studies used similar recruitment strategies [18, 38-40]. The final, and potentially most important, consideration is that subgroup analyses conducted on three of the papers detailed above showed that the most *complex* cases benefited more from a comprehensive (multi-component) program [27, 41, 42]. Complex cases were described as patients with a poor prognosis classification (from a screening instrument consisting of a combination of psychological, motivational, and physiotherapy factors) [41], depressed comorbidity [27], low job satisfaction, low work autonomy, no interest in returning to the same job, and those at risk of losing their job [42]. These findings on case complexity are in line with the interviews with professionals in Chapter 6, who acknowledged that these specific patient factors might guide treatment stratification.

Stepped-care approach

Through appraisal of the dosage and content arguments presented above, the practice of quasi-flexible and tailored VR can be justified. This is concurrent with the conclusions drawn from the interviews with patients, professionals, and managers in Chapter 6. It has been proposed that this approach is best instigated through simple, low-cost interventions, such as “brief interventions” (defined as a thorough examination by a physician, including reassurance and advice about staying active, with follow-up by a physiotherapist) [43]. The dosage (and content) is then increased for the most complex cases [43, 44]. This “stepped-care” approach is advocated in clinical guidelines [45, 46] and research [47, 48]. Waddel et al. [48] describe stepped care as an approach that is focused on the individual, allowing the allocation of resources to those most in need of

them and thus providing an effective framework for distributing resources. An important prerequisite for the stepped-care approach is the use of screening instruments that can stratify patients; for example, into low-, medium- and high-risk groups. There are some stratification instruments with proper clinimetric properties that have been developed for first-line use in pain care (the STarT back questionnaire and Östebro musculoskeletal pain screening questionnaire) [49-51]. However, such screening tools are not available for second- or third-line care, and these should be developed. Therefore, a goal for future research concerning the development of a stepped-care approach in VR would be to develop a screening tool which enables stratification of patients referred to VR. Some examples of this exist in the literature [41, 52-54], but these should be validated and studied in a Dutch VR setting.

Methodological considerations

In this paragraph, methodological considerations that are not explicitly addressed in Chapters 2-7 are discussed.

Clinimetric methodological considerations

One methodological consideration of the clinimetrics section of this thesis was the timing of the baseline questionnaire distribution. Patients received emails with login data and a request to complete the questionnaires online at baseline, discharge, and follow up. Baseline questionnaires are sent out before multidisciplinary screening is performed at the VR center. However, the time between the multidisciplinary screening and the start of VR was 8 ± 4.4 weeks (Chapter 4). A Swedish study [55] used the Örebro musculoskeletal pain screening questionnaire to classify patients with musculoskeletal pain into three subgroups: “medium risk,” “fear and avoidance,” and “emotional distress.” After 7 weeks—just prior to treatment—they repeated this procedure. They found that the subgroups classified at screening, typically seven weeks before treatment started, were not stable and that the probability that participants changed subgroup was high [55]. The authors of this study, therefore, recommended that “profiles and targets for interventions should be determined immediately prior to treatment start, preferably using full questionnaires (p. 518).” Since the baseline questionnaires described in Chapter 4 were completed, on average, 8 weeks before treatment began, it is likely that the results of the study were affected by this phenomenon (more specifically, the results that used baseline

and discharge data (Chapters 3, 4, and 7)). This evidence [55] should be used to improve data collection in clinical practice and allow for more meaningful patient stratification.

Another methodological consideration of the clinimetric element of this thesis was raised during the assessment of the iPCQ-VR (Chapter 3): whether the concept of "modified work" (sometimes described as "therapeutic work") should be included as a measure, and, if so, how to adequately achieve this. However, because workers are often unaware of when they are performing modified work, and because asking about this would violate criterion validity–patients do not understand the question–, it was decided to not add an extra question concerning modified work to the iPCQ-VR. Since the iPCQ-VR does not measure this feature, and since patients might classify modified work as real working hours in the "working status" and "working hours per week" items of the iPCQ-VR, these numbers are probably overrepresented. This might be particularly pronounced in the Netherlands, where modified work is an obligatory reintegration strategy for employers returning to work [56]. Some European countries also have a modified work program, but the majority of countries do not [35, 37]. The "working status" and "working hours per week" figures presented in Chapter 3 and Chapter 7 should, therefore, be interpreted within a Dutch context, and can not automatically be transferred to countries with different policies [35].

Dose-content methodological considerations

The design of a randomized controlled trial affords it the highest level of internal validity relative to other study designs; therefore, the RCT tends to be thought of as the highest level of evidence [57, 58]. However, there is a growing evidence base describing the disadvantages of RCTs, such as low external validity (caused by overly restrictive eligibility criteria), the large time-frames involved, and high costs. Moreover, it has been suggested that the RCT is not suitable for complex interventions such as VR [59-62]. VR can be described as "complex" because it consists of multiple components, providers, locations and outcomes, with varying degrees of interdependent from each other, and, therefore, interventions can be difficult to standardize or administer uniformly [61].

Nevertheless, in Chapter 5 a RCT was designed to evaluate the clinical and cost effectiveness of a 40-hour VR program compared to a 100-hour VR program, targeted at work participation in workers on sick leave due to CMP. Upon having to end the trial early (Figure 1), it became clear that this RCT was indeed

inappropriate in this Dutch VR setting. An observational study design was therefore used in Chapter 7. The observational design is lower in the levels of evidence hierarchy [58], because of a higher risk of bias compared to RCT design. However, advantages of the observational design compared with the RCT design is that it can investigate a broader range of exposures, has potentially greater generalizability, and tends to be less expensive [57].

In attempt to circumnavigate these problems in experimental and observational study design, several papers have provided alternatives for testing the efficacy of VR [57, 61, 62]. A promising observational design for use in the context of VR is the "propensity scores method," a statistical matching technique that can be applied to control for confounding in evaluative studies with observational data. The advantage of this design is that it mimics randomization through controlling for known prognostic factors and making groups homogeneous on baseline [61]. Another advantage is that it uses logistic regression, which simplifies interpretation. A disadvantage of the propensity score method is that very large sample sizes are needed [61]. Unfortunately, such large sample sizes were beyond the reach of this thesis. Nonetheless, two recent publications have shown the significant potential of this design [63, 64].

Another useful observational study design is the "interrupted time series design." In this design, a series of measurements are performed before and after the implementation of an intervention in order to detect whether the intervention has a significantly greater effect than the *underlying secular trend*, such as an economic, market, or demographic trends [61, 62, 65]. Advantages of this design are that randomization is not necessary and that routinely collected data can be used, such as workers' medical examinations, income insurance data, or workers' compensation data, which increases feasibility. A disadvantage is the determination of the time-frame and the quality and availability of the data (e.g., monthly or yearly data of sickness absence); which are required to detect correlations or trends before and after the implementation of a (new) approach or intervention [61]. However, this design was unavailable to this thesis, as it was not feasible to implement less-comprehensive VR programs in the participating centers. However, this could be a useful study design in the near future, if more centers will implement less-comprehensive VR programs as standard in the Netherlands.

Recommendations

This thesis has provided new insights and knowledge concerning the clinimetrics-related and dose-content research gaps identified in VR, contributing to the overall quality improvement of the discipline. To further improve the quality of VR, the following recommendations for VR professionals, practices, researchers, and future research are suggested:

Recommendations for VR professionals

- Professionals are encouraged to elicit the return to work expectations of the patient at baseline when creating an individually-tailored VR program. Since patients with low RTW expectations are three times less likely to achieve work participation at discharge and six-month follow-up, this group should receive specific attention.
- Establish those work-related components that have already been accomplished at baseline, or that can be expected to be accomplished over the patient's work. Use this information to decide, together with the patient, employer and occupational physician, whether an additional work module is justified.
- During the intervention period, ask patients at multiple points in time how useful specific treatment components are and, if necessary, act upon these insights.
- Link the PDI baseline score to the corresponding change scores as reported in Chapter 4, and perform responder analysis. This information can then be used for evaluative purposes at an individual patient level or can be used for benchmarking purposes at a group level.

Recommendations for VR practice

- It is recommended that VR practices in the Netherlands use the VR-pain Core Set for data collection. This will increase knowledge transfer [66] and fosters benchmarking.
- The three domains described by Cullen [67] (health, coordination, and work) can act as a starting point concerning the content of VR. A next step could be to personalize program content and dosage to the specific needs of the patient. Quasi-flexible and tailored VR could be applied. The operationalization of such an approach should be accompanied by future research.

Recommendations for VR researchers

- If patient-reported questionnaires are used for data collection, consider using the "working status" and "working hours per week" items of the iPCQ-VR as a proxy for the assessment of work participation.
- Consider using the responsiveness and change scores of the iPCQ-VR/PDI to perform "responder analyses" for efficacy study aims. However, it should be recommended that researchers are encouraged to calculate the responsiveness and change scores of their specific study population and context since the minimal important change values for instruments differs widely between studies [14, 68].
- Because RTW expectation is an important predictor of work participation at discharge and six-month follow-up, it is recommended that this measure is assessed at baseline and corrected for when analyzing the results from interventional programs on work participation, or when prognostic studies are conducted.

Recommendations for future research

- For research into cost effectiveness, consider all possible research designs exhaustively. Researchers are discouraged *automatically* opt for a RCT design and are encouraged to consider alternative study designs (e.g., experimental and observational) [61].
- Since multidisciplinary programs are standard practice in the treatment of patients with CMP and declined work participation in most industrialized countries, it is not recommended to use "care as usual" as the control group, which may include general practitioner-based care, occupational practitioner-based care, or other, monodisciplinary, first line care. Researchers should, instead, compare innovative multidisciplinary programs with the current standard (standard multidisciplinary practice).
- Develop the iPCQ-VR and TiCP-VR questionnaires further, as proposed earlier in this chapter.
- Expand on the research and design of a stepped-care VR approach. An initial suggestion would be to develop and validate a stratification instrument which can be used in VR context.

Epilogue

This thesis aimed to contribute to the quality improvement of VR. This aim has been accomplished in several ways. First, a core set of diagnostic and evaluative measures specifically designed for use in Dutch VR centers has been developed. This enhances benchmarking between centers, other patient groups, and the scientific literature. Second, the clinimetric properties of relevant questionnaires were examined in a Dutch VR context. This provided information detailing which instruments, questionnaire items, and cut-off scores can be used for diagnostic, process-related, and evaluative purposes in VR clinical practice and research. Third, the experiences of patients, professionals, and managers with comprehensive (standard practice) and less-comprehensive (experimental) VR programs were collected. These provided insights that can be used both to develop new programs and to refine existing VR programs. Finally, describing the relationship between work participation and multicomponent VR programs both *with* and *without* an additional work module has allowed for insights that can influence VR program content choice and facilitate patient stratification.

With this thesis, several steps are made that contribute to quality improvement in vocational rehabilitation. In order to further improve quality, a number of recommendations for practice and research are provided.

References

1. Young AE, Viikari-Juntura E, Boot CR, Chan C, Gimeno Ruiz de Porras D, Linton SJ. Workplace Outcomes in Work-Disability Prevention Research: A Review with Recommendations for Future Research. J Occup Rehabil. 2016 Dec;26(4):434-47.
2. Sabariego C, Coenen M, Ito E, Fheodoroff K, Scaratti C, Leonardi M, et al. Effectiveness of Integration and Re-Integration into Work Strategies for Persons with Chronic Conditions: A Systematic Review of European Strategies. Int J Environ Res Public Health. 2018 Mar 19;15(3).
3. Steenstra IA, Lee H, de Vroome EM, Busse JW, Hogg-Johnson SJ. Comparing current definitions of return to work: a measurement approach. J Occup Rehabil. 2012 Sep;22(3):394-400.
4. van Vilsteren M, van Oostrom SH, de Vet HC, Franche RL, Boot CR, Anema JR. Workplace interventions to prevent work disability in workers on sick leave. Cochrane Database Syst Rev. 2015 Oct 5(10):CD006955.
5. van Dongen JM, van Wier MF, Tompa E, Bongers PM, van der Beek AJ, van Tulder MW, et al. Trial-based economic evaluations in occupational health: principles, methods, and recommendations. J Occup Environ Med. 2014 Jun;56(6):563-72.
6. Zhang W, Bansback N, Anis AH. Measuring and valuing productivity loss due to poor health: A critical review. Soc Sci Med. 2011 Jan;72(2):185-92.
7. Leggett S, van der Zee-Neuen A, Boonen A, Beaton DE, Bojinca M, Bosworth A, et al. Test-retest Reliability and Correlations of 5 Global Measures Addressing At-work Productivity Loss in Patients with Rheumatic Diseases. J Rheumatol. 2016 Feb;43(2):433-9.
8. Gardner BT, Dale AM, Buckner-Petty S, Van Dillen L, Amick BC, 3rd, Evanoff B. Comparison of Employer Productivity Metrics to Lost Productivity Estimated by Commonly Used Questionnaires. J Occup Environ Med. 2016 Feb;58(2):170-7.
9. Tang K. Estimating productivity costs in health economic evaluations: a review of instruments and psychometric evidence. Pharmacoeconomics. 2015 Jan;33(1):31-48.
10. Boschman JS, Nieuwenhuijsen K, Sluiter JK. Within-person fluctuations in wellbeing and task-specific work ability. Qual Life Res. 2018 Feb;27(2):437-46.
11. Bouwmans C, Hakkaart- van Roijen L, Koopmanschap M, Krol M, H Severens H, W Brouwer W. iMTA Medical Consumption Questionnaire [Dutch manual]. Rotterdam: Institute for Medical Technology Assessment (iMTA) 2013.
12. Hakkaart-van Roijen L, van der Linden N, Bouwmans C, Kanters T, Tan S. Cost guide: Methodology of cost research and reference prices for economic evaluations in health care [In Dutch: Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg]: Institute for Medical Technology Assessment (iMTA) 2015.
13. Baker DW, Hays RD, Brook RH. Understanding changes in health status. Is the floor phenomenon merely the last step of the staircase? Med Care. 1997 Jan;35(1):1-15.
14. Olsen MF, Bjerre E, Hansen MD, Tendal B, Hilden J, Hrobjartsson A. Minimum clinically important differences in chronic pain vary considerably by baseline pain and methodological factors: systematic review of empirical studies. J Clin Epidemiol. 2018 Sep;101:87-106.

15. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Ann Rheum Dis. 2005 Jan;64(1):29-33.

16. Aasdahl L, Pape K, Vasseljen O, Johnsen R, Gismervik S, Halsteinli V, et al. Effect of Inpatient Multicomponent Occupational Rehabilitation Versus Less Comprehensive Outpatient Rehabilitation on Sickness Absence in Persons with Musculoskeletal- or Mental Health Disorders: A Randomized Clinical Trial. J Occup Rehabil. 2017 Apr 11.

17. Brendbekken R, Eriksen HR, Grasdal A, Harris A, Hagen EM, Tangen T. Return to Work in Patients with Chronic Musculoskeletal Pain: Multidisciplinary Intervention Versus Brief Intervention: A Randomized Clinical Trial. J Occup Rehabil. 2017 Mar;27(1):82-91.

18. Moll LT, Jensen OK, Schiottz-Christensen B, Stapelfeldt CM, Christiansen DH, Nielsen CV, et al. Return to Work in Employees on Sick Leave due to Neck or Shoulder Pain: A Randomized Clinical Trial Comparing Multidisciplinary and Brief Intervention with One-Year Register-Based Follow-Up. J Occup Rehabil. 2018 Jun;28(2):346-56.

19. Ronzi Y, Roche-Leboucher G, Begue C, Dubus V, Bontoux L, Roquelaure Y, et al. Efficiency of three treatment strategies on occupational and quality of life impairments for chronic low back pain patients: is the multidisciplinary approach the key feature to success? Clin Rehabil. 2017 Oct;31(10):1364-73.

20. Jensen C, Jensen OK, Nielsen CV. Sustainability of return to work in sick-listed employees with low-back pain. Two-year follow-up in a randomized clinical trial comparing multidisciplinary and brief intervention. BMC Musculoskelet Disord. 2012 Aug 25;13:156.

21. Myhre K, Marchand GH, Leivseth G, Keller A, Bautz-Holter E, Sandvik L, et al. The effect of work-focused rehabilitation among patients with neck and back pain: a randomized controlled trial. Spine (Phila Pa 1976). 2014 Nov 15;39(24):1999-2006.

22. Reme SE, Tveito TH, Harris A, Lie SA, Grasdal A, Indahl A, et al. Cognitive Interventions and Nutritional Supplements (The CINS Trial): A Randomized Controlled, Multicenter Trial Comparing a Brief Intervention With Additional Cognitive Behavioral Therapy, Seal Oil, and Soy Oil for Sick-Listed Low Back Pain Patients. Spine (Phila Pa 1976). 2016 Oct 15;41(20):1557-64.

23. Falkenstrom F, Josefsson A, Berggren T, Holmqvist R. How much therapy is enough? Comparing dose-effect and good-enough models in two different settings. Psychotherapy (Chic). 2016 Mar;53(1):130-9.

24. Owen JJ, Adelson J, Budge S, Kopta SM, Reese RJ. Good-enough level and dose-effect models: Variation among outcomes and therapists. Psychother Res. 2016;26(1):22-30.

25. Baldwin SA, Berkeljon A, Atkins DC, Olsen JA, Nielsen SL. Rates of change in naturalistic psychotherapy: contrasting dose-effect and good-enough level models of change. J Consult Clin Psychol. 2009 Apr;77(2):203-11.

26. Reneman MF, Waterschoot FPC, Bennen E, Schiphorst Preuper HR, Dijkstra PU, Geertzen JHB. Dosage of pain rehabilitation programs: a qualitative study from patient and professionals' perspectives. BMC Musculoskelet Disord. 2018 Jun 30;19(1):206.

27. Harris A, Moe TF, Eriksen HR, Tangen T, Lie SA, Tveito TH, et al. Brief intervention, physical exercise and cognitive behavioural group therapy for patients with chronic low back pain (The CINS trial). Eur J Pain. 2017 Sep;21(8):1397-407.

28. Lambeek LC, Bosmans JE, Van Royen BJ, Van Tulder MW, Van Mechelen W, Anema JR. Effect of integrated care for sick listed patients with chronic low back pain: economic evaluation alongside a randomised controlled trial. BMJ. 2010;341:c6414.

29. Bair MJ, Matthias MS, Nyland KA, Huffman MA, Stubbs DL, Kroenke K, et al. Barriers and facilitators to chronic pain self-management: a qualitative study of primary care patients with comorbid musculoskeletal pain and depression. Pain Med. 2009 Oct;10(7):1280-90.

30. Carnes D, Homer KE, Miles CL, Pincus T, Underwood M, Rahman A, et al. Effective delivery styles and content for self-management interventions for chronic musculoskeletal pain: a systematic literature review. Clin J Pain. 2012 May;28(4):344-54.

31. Hutting N, Detaille SI, Heerkens YF, Engels JA, Staal JB, Nijhuis-van der Sanden MW. Experiences of Participants in a Self-Management Program for Employees with Complaints of the Arm, Neck or Shoulder (CANS): A Mixed Methods Study. J Occup Rehabil. 2017 Mar;27(1):35-48.

32. Jansson I, Perseius KI, Gunnarsson AB, Bjorklund A. Work and everyday activities: experiences from two interventions addressing people with common mental disorders. Scand J Occup Ther. 2014 Jul;21(4):295-304.

33. Steihaug S, Ahlsen B, Malterud K. From exercise and education to movement and interaction. Treatment groups in primary care for women with chronic muscular pain. Scand J Prim Health Care. 2001 Dec;19(4):249-54.

34. Steihaug S, Ahlsen B, Malterud K. "I am allowed to be myself": women with chronic muscular pain being recognized. Scand J Public Health. 2002;30(4):281-7.

35. Anema JR, Schellart AJ, Cassidy JD, Loisel P, Veerman TJ, van der Beek AJ. 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 Dec;19(4):419-26.

36. MacEachen E, Varatharajan S, Du B, Bartel E, Ekberg K. The Uneven Foci of Work Disability Research Across Cause-based and Comprehensive Social Security Systems. Int J Health Serv. 2018 Nov 14:20731418809857.

37. Sickness, Disability and Work: Breaking the Barriers. A synthesis of findings across OECD countries. OECD 2010.

38. Jensen C, Jensen OK, Christiansen DH, Nielsen CV. One-year follow-up in employees sick-listed because of low back pain: randomized clinical trial comparing multidisciplinary and brief intervention. Spine (Phila Pa 1976). 2011 Jul 1;36(15):1180-9.

39. Lambeek LC, Anema JR, van Royen BJ, Buijs PC, Wuisman PI, van Tulder MW, et al. Multidisciplinary outpatient care program for patients with chronic low back pain: design of a randomized controlled trial and cost-effectiveness study [ISRCTN28478651]. BMC Public Health. 2007;7:254.

40. Fimland MS, Vasseljen O, Gismervik S, Rise MB, Halsteinli V, Jacobsen HB, et al. Occupational rehabilitation programs for musculoskeletal pain and common mental health disorders: study protocol of a randomized controlled trial. BMC Public Health. 2014 Apr 16;14:368.

41. Haldorsen EM, Grasdal AL, Skouen JS, Risa AE, Kronholm K, Ursin H. Is there a right treatment for a particular patient group? Comparison of ordinary treatment, light multidisciplinary treatment, and extensive multidisciplinary treatment for long-term sick-listed employees with musculoskeletal pain. Pain. 2002 Jan;95(1-2):49-63.
42. Stapelfeldt CM, Christiansen DH, Jensen OK, Nielsen CV, Petersen KD, Jensen C. Subgroup analyses on return to work in sick-listed employees with low back pain in a randomised trial comparing brief and multidisciplinary intervention. BMC Musculoskelet Disord. 2011 May 25;12:112.
43. Aasdahl L, Fimland MS. Is there really a "golden hour" for work disability interventions? a narrative review. Disabil Rehabil. 2019 Jan 3:1-8.
44. Palmer KT, Harris EC, Linaker C, Barker M, Lawrence W, Cooper C, et al. Effectiveness of community- and workplace-based interventions to manage musculoskeletal-related sickness absence and job loss: a systematic review. Rheumatology (Oxford). 2012 Feb;51(2):230-42.
45. NHG-Werkgroep pijn. NHG-Standaard Pijn (Dutch General Practitioners Society-Standard Pain). Huisarts Wet. 2015;58(9):472-85.
46. Zorgstandaard Chronische Pijn (Care Standard Chronic Pain). PA!N 2017.
47. Cochrane A, Higgins NM, FitzGerald O, Gallagher P, Ashton J, Corcoran O, et al. Early interventions to promote work participation in people with regional musculoskeletal pain: a systematic review and meta-analysis. Clin Rehabil. 2017 Nov;31(11):1466-81.
48. Waddell G, Burton A, Kendall N. Vocational rehabilitation: what works, for whom, and when? The Stationery Office 2013.
49. Hill JC, Whitehurst DG, Lewis M, Bryan S, Dunn KM, Foster NE, et al. Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. Lancet. 2011 Oct 29;378(9802):1560-71.
50. Linton SJ, Boersma K, Traczyk M, Shaw W, Nicholas M. Early Workplace Communication and Problem Solving to Prevent Back Disability: Results of a Randomized Controlled Trial Among High-Risk Workers and Their Supervisors. J Occup Rehabil. 2016 Jun;26(2):150-9.
51. Linton SJ, Nicholas M, MacDonald S. Development of a short form of the Orebro Musculoskeletal Pain Screening Questionnaire. Spine (Phila Pa 1976). 2011 Oct 15;36(22):1891-5.
52. Core Standards for Pain Management Services in the UK. CSPMS UK 2015 (p. 121).
53. Waterschoot FP, Bennen E, van der Woude LH, Schiphorst Preuper HR, Reneman MF. Case complexity in patients with chronic nonspecific musculoskeletal pain: a Delphi and feasibility study. Int J Rehabil Res. 2016 Mar;39(1):48-56.
54. Pain Management Programs – Which Patient for Which Program? Agency for Clinical Innovation 2013.
55. Bergbom S, Boersma K, Linton SJ. When Matching Fails: Understanding the Process of Matching Pain-Disability Treatment to Risk Profile. J Occup Rehabil. 2015 Sep;25(3):518-26.
56. Wet verbetering poortwachter (The Law on Gate keeping Disability Insurance). Accessed 22 January 2018; Available from: https://www.arboportaal.nl/onderwerpen/wet-verbetering-poortwachter.

57. Golfam M, Beall R, Brehaut J, Saeed S, Relton C, Ashbury FD, et al. Comparing alternative design options for chronic disease prevention interventions. Eur J Clin Invest. 2015 Jan;45(1):87-99.

58. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. Lancet. 2017 Jul 22;390(10092):415-23.

59. Bothwell LE, Greene JA, Podolsky SH, Jones DS. Assessing the Gold Standard--Lessons from the History of RCTs. N Engl J Med. 2016 Jun 2;374(22):2175-81.

60. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Int J Nurs Stud. 2013 May;50(5):587-92.

61. Schelvis RM, Oude Hengel KM, Burdorf A, Blatter BM, Strijk JE, van der Beek AJ. Evaluation of occupational health interventions using a randomized controlled trial: challenges and alternative research designs. Scand J Work Environ Health. 2015 Sep 1;41(5):491-503.

62. Williams-Whitt K, Bultmann U, Amick B, 3rd, Munir F, Tveito TH, Anema JR. Workplace Interventions to Prevent Disability from Both the Scientific and Practice Perspectives: A Comparison of Scientific Literature, Grey Literature and Stakeholder Observations. J Occup Rehabil. 2016 Dec;26(4):417-33.

63. Streibelt M, Burger W, Nieuwenhuijsen K, Bethge M. Effectiveness of Graded Return to Work After Multimodal Rehabilitation in Patients with Mental Disorders: A Propensity Score Analysis. J Occup Rehabil. 2018 Mar;28(1):180-9.

64. Viikari-Juntura E, Virta LJ, Kausto J, Autti-Ramo I, Martimo KP, Laaksonen M, et al. Legislative change enabling use of early part-time sick leave enhanced return to work and work participation in Finland. Scand J Work Environ Health. 2017 Sep 1;43(5):447-56.

65. Fine B, Schultz SE, White L, Henry D. Impact of restricting diagnostic imaging reimbursement for uncomplicated low back pain in Ontario: a population-based interrupted time series analysis. CMAJ Open. 2017 Oct 13;5(4).

66. Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for patients. Trials. 2017 Mar 14;18(1):122.

67. Cullen KL, Irvin E, Collie A, Clay F, Gensby U, Jennings PA, et al. Effectiveness of Workplace Interventions in Return-to-Work for Musculoskeletal, Pain-Related and Mental Health Conditions: An Update of the Evidence and Messages for Practitioners. J Occup Rehabil. 2018 Mar;28(1):1-15.

68. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. J Clin Epidemiol. 2010 May;63(5):524-34.

CHAPTER 9

Summary

Chronic musculoskeletal pain (CMP) has a negative impact on a person's general functioning and is associated with high levels of productivity loss from work and significant socioeconomic impact. Achieving sustainable levels of work participation by workers with CMP is of significant importance from both societal and individual perspectives.

Research has shown that biopsychosocial multi-domain vocational rehabilitation (VR) is beneficial in achieving sustainable levels of work participation for sick-listed workers with CMP. VR can be understood as an interdisciplinary, multi-domain intervention program, comprising multimodal treatments provided by a multidisciplinary team, collaborating in the assessment and treatment of patients using a biopsychosocial model and shared goals.

There are however a number of research gaps concerning "clinimetrics" and "dose-content" in VR. In this thesis, clinimetrics is looked at in relation to the absence of a "VR-pain Core Set" of questionnaires measuring the concepts of pain and work together, and the measurement properties of meaningful questionnaires, such as "work participation", "pain-related disability", and "healthcare usage." Dose-content is looked at in relation to the gap in knowledge about the duration, frequency, contact hours, and content of VR programs applied in research and common practice. These research gaps were examined in this thesis. The overall aim of this thesis was to contribute to the quality improvement of VR for patients with chronic musculoskeletal pain and reduced work participation.

The aims of this thesis were divided into two parts:

I. To investigate the clinimetric properties of work participation, healthcare usage, and pain-related disability measures;

II. To investigate the relationship between the dosage and content of VR on work participation.

Four research questions were formulated to address the aims of this thesis.

Research question 1: *Which questionnaires should be included in a focused "VR-pain Core Set" that can be used across VR practice in the Netherlands and can examine clinical and cost effectiveness?*

A consensus-based core set of diagnostic and evaluative measures— specifically tailored for use in the context of Dutch VR centers for the VR of patients with subacute and chronic musculoskeletal pain— was described in Chapter 2. To accomplish this "VR-pain Core Set", the brief ICF (International Classification of Functioning, Disability and Health) core set for VR was used as the reference framework in VR, and IMMPACT (the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials) recommendations were used in the outcome measurements around pain. These two existing outcome sets were merged, and irrelevant items were deleted. Next, the remaining domains were matched with existing instruments or measures. In a final step, a preliminary core set was judged by proposed users (VR clinicians), Dutch pain rehabilitation experts, and international VR experts. With this range of input, a final core set was developed consisting of 18 relevant domains for pain and VR and 12 validated instruments measuring these domains.

The developed VR-pain Core Set consists of items from the following questionnaire tools: EuroQol 5 Dimensions (EQ-5D), Work Ability Index (WAI), productivity and disease questionnaire [PRODISQ, later replaced by the iMTA productivity Cost Questionnaire-Vocational Rehabilitation (iPCQ-VR)], Pain Disability Index (PDI), RAND-36 physical functioning scale, work reintegration questionnaire (WRQ), Numeric Rating Scale (NRS) pain, NRS fatigue, lifting test, Astrand bicycle test or Bruce treadmill test, Trimbos iMTA questionnaire for measuring costs of psychiatric illnesses (TiC-P, later replaced by the TiCP-VR), and the Global Perceived Effect (GPE). Of these, iPCQ-VR, TiCP-VR, and EQ-5D can be used for cost-effectiveness purposes.

After the development of the VR-pain Core Set, it was adapted to a web-based input structure and adopted by seven Dutch VR centers. From 2014 until today, this core set is used to collect data for diagnostic, evaluative, and scientific purposes. In clinical practice, patients complete web-based questionnaires at baseline, discharge, and at six and twelve-month follow-up.

Research question 2*: What are the clinimetric properties of work participation, healthcare usage, and pain-related disability questionnaires for patients with CMP and reduced work participation in attendance of, and following discharge from VR, in the Netherlands?*

In chapter 3, test-retest reliability, agreement, and patient responsiveness to the iPCQ-VR questionnaire, which measures work participation, were assessed.

Test-retest reliability and agreement were assessed with a two-week interval. Responsiveness was assessed in VR treatment, and following discharge from a VR program of 15 weeks duration. Data was obtained from six VR centers in the Netherlands. Test-retest reliability was determined with an intraclass correlation coefficient (ICC) and Cohen's kappa (k). Agreement was determined by the Standard Error of Measurement (SEM), smallest detectable changes (on group and individual level), and percentage observed, positive and negative agreement. Responsiveness was determined with area under the curve (AUC) obtained from a receiver operating characteristic curve (ROC-curve). Agreement was studied on 50 participants, retest reliability on 16-23 *stable* participants, and responsiveness was studied on 223 participants. The iPCQ-VR showed good measurement properties on "working status" (k 0.96), "number of hours working per week" (ICC 0.90; AUC 0.86[1]), and "long-term sick leave" (k 0.74). Low measurement properties were found for "short-term sick leave" (k 0.45; ICC 0.54; AUC 0.66[4]) and "presenteeism" [k 0.42; ICC 0.52 (days), 0.56 (score); AUC 0.55 (days), 0.60 (score)].

The retest reliability and agreement of the TiCP-VR questionnaire, which measures health care usage, were also examined in Chapter 3. The TiCP-VR showed adequate reliability on "all healthcare utilization items together" (ICC 0.81) and "medication use" (k 0.78), but showed low agreement (e.g. low kappa scores) on the single healthcare utilization items.

In chapter 4, responsiveness and interpreting the change scores concerning the Pain Disability Index (PDI) questionnaire were examined. Retrospective data from patients with CMP that underwent vocational rehabilitation between 2014 and 2017 was used. The anchor-based method was used to assess the responsiveness of the PDI in VR treatment, and following discharge from a VR program, of 15 weeks duration. The responsiveness was also examined by looking at subsamples based on PDI baseline quartile scores. An ROC-curve was performed, including Area Under the Curve (AUC) and Minimal Important Change (MIC). Analyses were performed on a total sample of 341 participants and on subsamples based on PDI baseline quartile scores. The results showed that the PDI was responsive to detect real changes in pain-related disability after VR treatment (AUC 0.79). A PDI change score of 13 points (MIC 12.5) can be considered as a real change in pain-related disability for the total study sample, and a PDI change score of

[4] This AUC was found on a subgroup of patients who were at full sick leave at baseline.

7-20 points can be considered as a real change in pain-related disability for the lowest and highest subsamples based on PDI baseline scores.

In summary, from Chapter 4 it can be concluded that the PDI can detect real change in patients with CMP upon discharge of VR. To interpret a PDI change score at the discharge of VR as "real change", patients with a PDI baseline score of ≤27 should decrease a minimum of 7 points; patients with a baseline score between 28 and 42 should decrease a minimum of 15 points; and patients with a baseline score ≥43 should decrease a minimum of 20 points. These results can be used in clinical practice and research to perform "responder analysis". Responder analysis focuses on the percentage of patients that have reached real change. This has been recommended as a readily interpretable measure that can be of relevance for both researchers and clinicians.

Research question 3: *What are the opinions and experiences of patients, professionals, and managers regarding the usefulness and feasibility of "comprehensive" and "less-comprehensive" VR programs?*

Chapter 5 contains a study protocol paper of a multicenter randomized controlled trial (RCT). The aim of the paper was to describe the design of a non-inferiority trial evaluating the effectiveness and cost-effectiveness of 40-hour "less comprehensive" VR (LC-VR) compared with 100-hour "comprehensive" VR (C-VR) on work participation for workers on sick leave due to subacute or chronic musculoskeletal pain. The RCT was conducted between 2014-2016, but was discontinued because too few participants were included in the study.

In Chapter 6, the purpose of the study was to explore the usefulness and feasibility of the C-VR program and the LC-VR program for workers on sick leave due to CMP. Semi-structured interviews using topic lists were held with seven patients placed into the LC-VR program, six patients in the C-VR program, and with professionals (n=8) and managers (n=9). All interviews were transcribed verbatim. Data was analyzed with systematic text condensation using thematic analysis. Three themes emerged for usefulness ("patient factors", "content", and "dosage") and six themes emerged for feasibility ("satisfaction", "intention to continue use", "perceived appropriateness", "positive/negative effects on target participants", "factors affecting implementation ease or difficulty", and "adaptations").

The main findings include, from the patients point of view, that both programs are considered feasible and generally useful. However, some patients stated that not all of the content was useful, and in some cases, content saturation took place. Professionals preferred working with the "C-VR as standard" program, although some disliked its rigid and uniform character. Professionals also felt that the C-VR program was too extensive for some patients and that these patients would likely benefit from the LC-VR program. Several patient factors were identified by professionals that might help stratification of patients into either the C-VR or LC-VR program. Managers felt that, despite appreciating the relevance of the LC-VR program, implementation of the program would not be financially possible due to the Dutch healthcare system. The overall conclusion from the patients, professionals, and managers was that it is not useful to deliver one VR program for all patients and that treatment should be personalized through the use of quasi-flexible and tailored VR programming.

Research question 4: *Are patients with CMP and reduced work participation who attended "VR with work module" more likely to achieve work participation than patients who attended "VR without work module?"*

In Chapter 7, a retrospective cohort study looks at the relationship between VR —with and without an additional work module— on the work participation of patients with CMP and reduced work participation, both following discharge from a VR program of 15 weeks duration, and follow-up six months later. Retrospective data was retrieved from care as usual provided by seven VR centers in the Netherlands. The VR program without work module ("VR") consisted of multi-component health care (physical exercise, cognitive behavioral therapy, education, relaxation). The VR program with an additional work module ("VR+") consisted of additional case management and workplace visit components. A multivariate logistic regression model was applied. The dependent variable was work participation (achieved/not achieved). Independent variables were type of intervention (VR/VR+), demographics, clinical, and work-related (return to work [RTW] expectation, sick leave duration, working status, job strain, and job dissatisfaction). The results showed that of the 142 patients included, 26% received the VR program and 74% received the VR+ program. Both programs increased work participation (proportion full and part-time work) at discharge (VR 71%; VR+ 80%) and at six-month follow-up (VR 80%, VR+ 86%). However, there were non-significant relationships between the type of intervention and work participation on discharge (OR 1.0, $p = 0.99$) and six-month follow-up (OR 1.3, $p = 0.52$). The RTW expectation was the only significant independent factor

in the multivariate model on discharge (OR 2.9, p = 0.00) and six-month follow-up (OR 3.0, p = 0.00). This means that patients who are positive about there return to work within six months from baseline have three times more chance of achieving successful work participation on discharge and six-months follow-up.

In conclusion, both programs led to increased work participation at discharge and six-month follow-up, but the addition of a work module to the VR program did not increase work participation.

Conclusion and recommendations for further research and practice

This thesis contributes to understanding how to improve the quality of VR. This has been achieved in several ways. First, a core set of diagnostic and evaluative measures specifically designed for use in Dutch VR centers has been developed. It is recommended that VR practices in the Netherlands use the VR-pain Core Set for data collection. This will increase knowledge transfer and enables benchmarking.

Second, the clinimetric properties of relevant questionnaires were examined in a Dutch VR context. This provided information detailing which instruments, questionnaire items, and cut-off scores can be used for diagnostic, process-related, and evaluative purposes in VR clinical practice and research. For instance, professionals and researchers might consider using the responsiveness and change scores of the iPCQ-VR/PDI questionnaires to perform responder analyses. This information can be used for evaluative purposes at individual patient level or can be used for benchmarking purposes at group level.

Third, the experiences of patients, professionals, and managers with C-VR and LC-VR programs were collected. The overall conclusion is that it is not useful to deliver one VR program for all patients and that treatment should be personalized through the use of quasi-flexible and tailored VR programming. Hence, future research might focus on the design and benefits of a stepped-care VR approach. An initial suggestion would be to develop and validate a stratification instrument which can be used in a VR context.

Finally, since the relationship between multicomponent VR both *with* and *without* an additional work module on work participation was non-significant, it should be recommended to not include a work module to VR as part of standardized care, but the decision to include or not should be determined on a patient by patient basis.

CHAPTER 10

Samenvatting

Mensen met chronische pijn aan het bewegingsapparaat (in dit proefschrift afgekort tot CMP[5]) worden beperkt in het uitvoeren van dagelijkse activiteiten, hebben een verminderde arbeidsparticipatie (verzuim en productiviteitsverlies in het werk) en hebben relatief veel medische zorg nodig. Dit heeft hoge maatschappelijke kosten tot gevolg; de verminderde arbeidsparticipatie beslaat ongeveer 80% van die kosten. Het is daarom vanuit individueel en maatschappelijk perspectief van groot belang om duurzame arbeidsparticipatie bij mensen met CMP te bevorderen.

Uit onderzoek blijkt dat het mogelijk is om door middel van arbeidsrevalidatie duurzame arbeidsparticipatie te bewerkstelligen en arbeidsverzuim te verminderen. Arbeidsrevalidatie bestaat over het algemeen uit diverse bio-psychosociale behandelmodules. De behandeling wordt veelal uitgevoerd door een multidisciplinair team dat interdisciplinair samenwerkt bij de triage en behandeling van patiënten met CMP en arbeidsverzuim. Uit de wetenschappelijke literatuur komen echter twee lacunes naar voren. De eerste lacune heeft betrekking op de “klinimetrie” van meetinstrumenten die binnen de arbeidsrevalidatie worden ingezet. Als lacune op het gebied van klinimetrie wordt in dit proefschrift bedoeld de afwezigheid van een “core set” van vragenlijsten -die de concepten arbeid en pijn gezamenlijk meten en evalueren- en de afwezigheid van relevante meeteigenschappen van belangrijke vragenlijsten binnen de arbeidsrevalidatie; voorbeelden hiervan zijn vragenlijsten voor het meten van “arbeidsparticipatie”, “beperkingen in dagelijks leven” en “zorggebruik”. De tweede lacune heeft betrekking op de afwezigheid van kennis over de duur, frequentie, contacturen (dosis) en inhoud van arbeidsrevalidatie zoals deze wordt beschreven in de wetenschappelijke literatuur en zoals deze wordt uitgevoerd in de klinische praktijk van de arbeidsrevalidatie. Onderdelen van de twee genoemde lacunes zijn onderzocht en in dit proefschrift beschreven.

Het algemene doel van dit proefschrift is om bij te dragen aan een kwaliteitsverbetering van arbeidsrevalidatie voor mensen met CMP die een verminderde arbeidsparticipatie hebben.

[5] Chronische musculoskeletale pijn

Het proefschrift is opgebouwd uit twee delen:

I. Het onderzoeken van de klinimetrische eigenschappen van vragenlijsten op het gebied van arbeidsparticipatie, zorggebruik en beperkingen in dagelijks leven.

II. Onderzoeken hoe de hoeveelheid (dosis) en inhoud van arbeidsrevalidatie is gerelateerd aan arbeidsparticipatie.

De doelen van dit proefschrift zijn geoperationaliseerd in de volgende vier onderzoeksvragen.

Onderzoeksvraag 1: *Welke vragenlijsten moeten onderdeel uitmaken van een "core set" die kan worden gebruikt door arbeidsrevalidatiecentra in Nederland en die ingezet kan worden om de (kosten)effectiviteit van de geleverde zorg te onderzoeken?*

De ontwikkeling en samenstelling van een op consensus gebaseerde core set van vragenlijsten die bestaat uit diagnostische en evaluatieve vragenlijsten is beschreven in hoofdstuk 2. De ontwikkelde core set bestaat uit onderdelen/domeinen die afkomstig zijn van twee bestaande core sets op het gebied van "arbeid" en "pijn". Voor het construct arbeid zijn onderdelen van de verkorte International Classification of Functioning, Disability and Health (ICF) core set gebruikt als referentiekader. Voor het construct pijn is het Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) raamwerk als referentiekader gebruikt. Onderdelen van de ICF en IMMPACT core sets zijn samengevoegd, waarbij irrelevante items zijn verwijderd. Vervolgens zijn de resterende domeinen gekoppeld aan bestaande Nederlandse vragenlijsten. In een laatste stap is een voorlopige set vragenlijsten beoordeeld door beoogde gebruikers (clinici uit de arbeidsrevalidatie) en Nederlandse en internationale experts op het gebied van pijnrevalidatie. Nadat alle meningen zijn verzameld, is een definitieve core set ontwikkeld, bestaande uit 12 gevalideerde vragenlijsten: EuroQol 5 Dimensions (EQ-5D), Work Ability Index (WAI), PROductivity and DISease Questionnaire (PRODISQ, later vervangen door de iMTA productivity Cost Questionnaire-Vocational Rehabilitation (iPCQ-VR)), Pain Disability Index (PDI), RAND-36 subschaal fysiek functioneren, vragenlijst arbeidsreintegratie (VAR), Numeric Rating Scale (NRS) pijn, NRS-vermoeidheid, tiltest, Astrand fietstest of Bruce loopband test, Trimbos iMTA questionnaire for measuring costs of psychiatric illnesses (TiC-P, later vervangen door de TiCP-VR) en Global

Perceived Effect (GPE). Van deze vragenlijsten kunnen de iPCQ-VR, TiCP-VR en EQ-5D gebruikt worden voor kosteneffectiviteitsonderzoek.

Na het gereedkomen ervan is de core set digitaal ter beschikking gesteld. Vanaf 2014 tot heden wordt de core set gebruikt door zeven arbeidsrevalidatiecentra om gegevens te verzamelen voor diagnostische, evaluatieve en wetenschappelijke doeleinden. In de klinische praktijk vullen patiënten de vragenlijsten via internet in voor aanvang, bij ontslag, en zes en twaalf maanden na ontslag uit de arbeidsrevalidatie.

Onderzoeksvraag 2: *Wat zijn de klinimetrische eigenschappen van vragenlijsten op het gebied van arbeidsparticipatie, zorggebruik en beperkingen in dagelijks leven voor aanvang en na ontslag uit de arbeidsrevalidatie in Nederland voor patiënten met CMP en een verminderde arbeidsparticipatie?*

De betrouwbaarheid en responsiviteit van de iPCQ-VR vragenlijst, die arbeidsparticipatie meet, is onderzocht in hoofdstuk 3. De betrouwbaarheid is onderzocht met een tijdsinterval van 2 weken. De responsiviteit is onderzocht voor aanvang en na ontslag na een arbeidsrevalidatieprogramma van 15 weken. Voor dit onderzoek is gebruik gemaakt van retrospectieve door patiënten zelf gerapporteerde data, afkomstig van zes arbeidsrevalidatiecentra in Nederland. De betrouwbaarheid is onderzocht met de intraclass correlation coefficient (ICC) en Cohen's kappa (k). De responsiviteit is onderzocht met de anker methode en de Area Under the Curve (AUC). De betrouwbaarheid is onderzocht bij 16-23 "stabiele" participanten (dat wil zeggen: mensen die geen verandering aangaven op het te meten construct tijdens het tweede meetmoment ten opzichte van het eerste meetmoment) en de responsiviteit is onderzocht bij 223 participanten. De betrouwbaarheid van de iPCQ-VR is hoog voor de items "werkstatus" (k=0.96), "aantal uren werken per week" (ICC=0.90) en "langdurig arbeidsverzuim" (k=0.74). De responsiviteit van het item "aantal uren werken per week" is hoog (AUC=0.86). Een lage betrouwbaarheid en responsiviteit is gevonden voor de items "kortdurend arbeidsverzuim" (ICC=0.54; AUC=0.66); "presenteïsme, aantal dagen in afgelopen 4 weken" (ICC=0.52, AUC=0.55); en "presenteïsme, score (0-10)" (ICC=0.56, AUC=0.60).

De betrouwbaarheid van de TiCP-VR-vragenlijst, die zorggebruik meet, is onderzocht in hoofdstuk 3. De zorggebruik items laten opgeteld een redelijke betrouwbaarheid zien (ICC=0.81), evenals "medicatiegebruik" (k=0.78). De afzonderlijke TiCP-VR items hebben een lage betrouwbaarheid.

De responsiviteit en de interpretatie van veranderscores op de beperkingen in het dagelijks leven vragenlijst -de Pain Disability Index (PDI)- is onderzocht in hoofdstuk 4. Voor dit onderzoek is gebruik gemaakt van retrospectieve door patiënten zelf gerapporteerde data, verzameld in de periode 2014-2017 en afkomstig van zeven arbeidsrevalidatiecentra in Nederland. De anker methode is gebruikt om de responsiviteit van de PDI te onderzoeken, voor aanvang en na ontslag na een arbeidsrevalidatie periode van 15 weken. De responsiviteit en veranderscores zijn berekend voor zowel de totale populatie als voor subgroepen, ingedeeld op basis van kwartielscores van de PDI bij de start van het onderzoek (baseline). Voor de analyses is een receiver operating characteristic (ROC) curve gebruikt, inclusief AUC en Minimal Important Change (MIC). Data van 341 participanten zijn gebruikt voor de analyses. De resultaten laten zien dat de PDI een goede responsiviteit heeft (AUC=0.79); dat wil zeggen dat de PDI goed in staat is om "echte" veranderingen in beperkingen in het dagelijks leven na het doorlopen van arbeidsrevalidatie aan te tonen. Een PDI-veranderscore van 13 punten kan worden beschouwd als een echte verandering in ervaren beperkingen in het dagelijks leven bij ontslag na arbeidsrevalidatie voor de totale studiepopulatie. Voor de subgroepen gebaseerd op de PDI baseline kwartiel scores zijn veranderscores tussen 7 en 20 punten gevonden. Om bij ontslag na arbeidsrevalidatie te kunnen spreken over een "echte verandering" ten aanzien van beperkingen in het dagelijks leven, moeten patiënten met een PDI-baseline score van ≤27 minimaal 7 punten lager op de PDI scoren, patiënten met een baseline score tussen 28 en 42 moeten minimaal 15 punten lager op de PDI scoren, en patiënten met een baseline score ≥43 moeten minimaal 20 punten lager op de PDI scoren. Deze afkapwaardes kunnen in de klinische praktijk en in wetenschappelijk onderzoek worden gebruikt om "responders" van "non-responders" te onderscheiden.

Onderzoeksvraag 3: *Wat zijn de meningen en ervaringen van patiënten, professionals en managers ten aanzien van het nut en de haalbaarheid van arbeidsrevalidatie met een duur van 100 uur (uitgebreid programma) en met een duur van 40 uur (minder uitgebreid programma)?*

Hoofdstuk 5 bevat een beschrijving van de opzet van een multicenter gerandomiseerde studie naar de effectiviteit en kosteneffectiviteit van een uitgebreid versus een minder uitgebreid arbeidsrevalidatieprogramma. De multicenter studie is uitgevoerd in de periode 2014-2016, maar is voortijdig beëindigd vanwege onvoldoende deelnemers.

Het doel van het onderzoek in hoofdstuk 6 is om het nut en de haalbaarheid van het uitgebreide en minder uitgebreide arbeidsrevalidatieprogramma voor werknemers met CMP en arbeidsverzuim te onderzoeken. Hiervoor zijn semi-gestructureerde interviews afgenomen met zeven patiënten die zijn toegewezen aan het minder uitgebreide programma, en met zes patiënten die zijn toegewezen aan het uitgebreide programma. Tevens zijn er acht interviews met professionals en negen interviews met managers afgenomen. Alle interviews zijn uitgeschreven en geanalyseerd door middel van systematische tekstcondensatie met behulp van thematische analyse. Drie thema's zijn naar voren gekomen voor nut ("patiëntfactoren", "inhoud", "dosering") en zes thema's voor haalbaarheid ("tevredenheid", "intentie om door te gaan met gebruik", "ingeschatte geschiktheid", "positieve/negatieve effecten voor patiënten", "factoren die van invloed zijn op de implementatie", en "aanpassingen").

Door de patiënten is aangegeven dat beide programma's voor hen haalbaar en in het algemeen nuttig zijn. Daarnaast hebben sommige patiënten aangegeven dat niet alle interventie onderdelen nuttig voor hen zijn, zoals ontspanning sessies, sessies met de psycholoog, groepseducatie, en voorgestelde werkplekaanpassingen. De professionals geven de voorkeur aan het werken met en uitvoeren van het uitgebreide programma. Sommige professionals geven echter wel aan dat ze dit programma te uniform vinden, en een deel van de professionals geeft aan dat sommige patiënten wellicht meer baat hebben bij het minder uitgebreide programma. Er zijn diverse patiënten karakteristieken/ factoren genoemd die mogelijk gebruikt kunnen worden om patiënten te stratificeren naar het uitgebreide of minder uitgebreide programma, zoals factoren op het gebied van intelligentie, gedrag, type klacht(en), mentaal/ cognitief functioneren en werk. De managers onderschrijven de relevantie van het minder uitgebreide programma, maar geven aan dat de implementatie van een dergelijk minder uitgebreid programma door de inrichting van het Nederlandse gezondheidszorgsysteem financieel niet haalbaar is. De algemene conclusie van de patiënten, professionals en managers samen is dat het niet nuttig is om één arbeidsrevalidatieprogramma voor alle patiënten aan te bieden, maar dat de behandeling gepersonaliseerd moet worden door middel van "quasi-flexibele" en op maat gemaakte arbeidsrevalidatie.

Onderzoeksvraag 4: *Hebben patiënten met CMP en een verminderde arbeidsparticipatie na het doorlopen van arbeidsrevalidatie aangevuld met een werkmodule meer kans op een hogere arbeidsparticipatie in vergelijking met patiënten die arbeidsrevalidatie zonder werkmodule hebben gevolgd?*

Hoofdstuk 7 bevat een retrospectieve cohortstudie waarin de relatie tussen arbeidsrevalidatie -met en zonder een aanvullende werkmodule- op de arbeidsparticipatie van patiënten met CMP en een verminderde arbeidsparticipatie, zowel bij ontslag als zes maanden na ontslag uit een arbeidsrevalidatieprogramma van 15 weken, is onderzocht. Hiervoor is gebruik gemaakt van retrospectieve door patiënten zelf gerapporteerde data, verzameld bij zeven Nederlandse arbeidsrevalidatiecentra. De arbeidsrevalidatie zonder werkmodule (aangeduid als AR) bestaat uit meerdere componenten (functionele fysieke training/ graded activity, cognitieve gedragstherapie, educatie en ontspanning). De arbeidsrevalidatie met aanvullende werkmodule (aangeduid als AR+) bestaat uit de AR componenten, aangevuld met werkgerichte coaching, casemanagement en een werkplekbezoek. Om de onderzoeksvraag te beantwoorden is een multivariaat logistisch regressiemodel gebruikt. De afhankelijke variabele is arbeidsparticipatie (verbeterd/gelijk gebleven, niet verbeterd). De onafhankelijke variabelen zijn type interventie (AR/AR+), demografische variabelen, klinische en werk gerelateerde variabelen (verwachting met betrekking tot terugkeer naar werk, arbeidsverzuim duur, werkstatus, werkdruk en ontevredenheid over het werk). Uit de resultaten blijkt dat van de 142 patiënten, 26% AR hebben gevolgd en 74% AR+. De resultaten van beide programma's laten een toename zien in de arbeidsparticipatie (fulltime en parttime werkzaam opgeteld) bij ontslag (AR+ 80%, AR 71%) en zes maanden na ontslag (AR 80%, AR+ 86%). De relatie tussen type interventie en arbeidsparticipatie bij ontslag (OR=1.0, p=0.99) en zes maanden na ontslag (OR=1.3, p=0.52) is echter niet significant. De verwachting met betrekking tot terugkeer naar werk bij de start van beide programma's is de enige significante onafhankelijke variabele in het finale multivariate model bij ontslag (OR=2.9, p=0.00) en zes maanden na ontslag (OR=3.0, p=0.00). Dit betekent dat mensen die een positieve terugkeer naar werk verwachting hebben op baseline de kans op een succesvolle arbeidsparticipatie op ontslag en zes maanden na ontslag drie keer groter is in vergelijking met mensen die een negatieve terugkeer naar werk verwachting hebben op baseline.

Samenvattend: beide programma's resulteren in een toename van arbeidsparticipatie bij ontslag en zes maanden na ontslag, maar de toevoeging van een werkmodule aan arbeidsrevalidatie verhoogt de arbeidsparticipatie niet significant in vergelijking met arbeidsrevalidatie zonder werkmodule.

Conclusie en aanbevelingen voor de praktijk en verder onderzoek

De algemene conclusie van dit proefschrift is dat een kwaliteitsverbetering ten aanzien van (de kennis over) arbeidsrevalidatie voor patiënten met CMP en een verminderde arbeidsparticipatie kan worden bewerkstelligd. Deze conclusie is opgebouwd uit verschillende pijlers, waarbij de belangrijkste vier hierna worden samengevat, en waarbij aanbevelingen voor de praktijk en verder onderzoek worden gegeven. Ten eerste is een core set bestaande uit diagnostische en evaluatieve vragenlijsten/instrumenten ontwikkeld specifiek voor de Nederlandse arbeidsrevalidatie. De core set wordt sinds 2014 tot op heden gebruikt door zeven Nederlandse arbeidsrevalidatiecentra; het is aan te bevelen dat meerdere arbeidsrevalidatiecentra in Nederland deze core set gaan gebruiken voor het verzamelen van diagnostische en evaluatieve gegevens. Dit zal de kennisoverdracht vergroten en benchmarking bevorderen.

Ten tweede zijn de klinimetrische eigenschappen van relevante vragenlijsten onderzocht in de Nederlandse arbeidsrevalidatie. Dit heeft informatie opgeleverd over welke (items van) vragenlijsten en/of welke afkapwaardes gebruikt kunnen worden voor diagnostische en evaluatieve doeleinden in de klinische praktijk en wetenschappelijk onderzoek. Professionals en wetenschappers kunnen overwegen om de afkapwaardes van de iPCQ-VR/PDI vragenlijsten te gebruiken om analyses uit te voeren naar wie binnen een populatie baat heeft gehad bij AR of juist niet. Deze informatie kan worden gebruikt voor evaluatiedoeleinden op individueel en groepsniveau en zal benchmarking bevorderen.

Ten derde zijn de ervaringen van patiënten, professionals en managers met een uitgebreid en minder uitgebreid arbeidsrevalidatieprogramma verzameld. De algemene conclusie uit dit onderzoek is dat het niet nuttig is om één arbeidsrevalidatieprogramma voor alle patiënten aan te bieden, maar dat de behandeling het beste gepersonaliseerd kan worden door het gebruik van quasi-flexibele en op maat gemaakte arbeidsrevalidatie. Een stapsgewijze of stepped-care aanpak is een mogelijke benadering hiervoor. Een dergelijke aanpak zal echter specifiek voor de arbeidsrevalidatie ontwikkeld en getoetst moeten worden. Een mogelijke eerste stap kan zijn om een stratificatie-instrument te ontwikkelen en te valideren die in de arbeidsrevalidatie gebruikt kan worden.

Tot slot wordt aanbevolen om niet standaard een werkmodule toe te voegen aan arbeidsrevalidatie, maar om per patiënt te beslissen of het toevoegen van

een werkmodule nuttig wordt geacht ter bevordering van de arbeidsparticipatie van mensen met CMP.

APPENDICES

Timo Beemster werd –als onderdeel van een drieling– geboren op 6 oktober 1985 te Hoorn, Nederland. In 2003 behaalde hij zijn havo-diploma aan het Martinuscollege te Grootebroek. Vervolgens studeerde hij oefentherapie Mensendieck aan de Hogeschool van Amsterdam, waar hij in 2007 voor slaagde. Daarna volgde hij de studie Gezondheidswetenschappen aan de Vrije Universiteit Amsterdam, waarvoor hij zijn Master of Science titel behaalde in 2011. Naast deze studie heeft Timo tussen 2007-2011 gewerkt bij revalidatiecentrum Heliomare als fysiek trainer. In 2011 is Timo gestart als junior onderzoeker bij de afdeling Research & Development van Heliomare. Medio 2012 is hij gestart met zijn promotieonderzoek, wat heeft geresulteerd in dit proefschrift. In 2016-2017 heeft Timo zijn promotieonderzoek anderhalf jaar gecombineerd met de functie als data analist voor de chronische pijn afdeling van Heliomare. Parallel aan zijn promotieonderzoek heeft Timo de opleiding tot Epidemioloog B gevolgd, welke hij na zijn promotie wil afronden. Na zijn promotie gaat Timo bij Heliomare aan de slag als senior onderzoeker. Naast zijn werk is Timo ruim tien jaar loop- en conditietrainer van handbalsters, waarvan de laatste jaren op Eredivisie niveau.

In dit dankwoord wil ik een aantal mensen bedanken, zonder wie ik dit proefschrift niet zou kunnen hebben geschreven.

Allereerst professionals en patiënten: bedankt voor jullie inzet en deelname aan de diverse onderzoeken.
Raad van bestuur Heliomare: bedankt dat jullie mijn promotieonderzoek altijd zijn blijven steunen.
Leescommissie en oppositie: bedankt voor de genomen tijd en moeite om mijn proefschrift te lezen en beoordelen, en bij de verdediging aanwezig te zijn.

Promotiecommissie
Michiel, Monique, Coen, Judith: de afgelopen jaren hebben we 136 keer met elkaar overlegd over mijn promotieonderzoek. Dit komt neer op gemiddeld 17 overlegmomenten per jaar met een SD van 4. Dat jullie aandeel essentieel is geweest bij de totstandkoming van dit proefschrift is evident. Heel erg bedankt dat jullie altijd zijn blijven geloven in mij en mijn onderzoek. Dankzij jullie ben ik de kritische onderzoeker die ik nu ben.

Michiel: in 2011 in een bruine kroeg in Den Bosch hebben we de eerste contacten gelegd. Je vroeg destijds aan mij of ik uiteindelijk in de aula wilde staan. Ik had op dat moment geen idee wat je hiermee bedoelde. Natuurlijk doelde je op een promotietraject. Na een traject vol hindernissen is het mij gelukt. Je hebt me hierin altijd gesteund, zowel op inhoud als op persoonlijk vlak. Je kennis en kunde, humor, Groningse gezegden, en je oprechte interesse in zaken ook buiten het werk om heb ik zeer gewaardeerd. Bedankt!

Monique: tijdens een van onze eerste overleggen gaf ik aan: “Oké, ik ga het proberen.” Jij corrigeerde dit in een: “Nee, je gaat het doen.” Dit zinnetje heb ik gedurende mijn promotietraject ingeprent. Tijdens overleg momenten ben je altijd recht voor zijn raap, eerlijk, en niet snel tevreden. Ik heb veel van je geleerd en ben daardoor gegroeid als onderzoeker. Je deur stond altijd voor mij open, ook voor niet werk-gerelateerde zaken. Bedankt!

Coen: als jij niks in mij of mijn onderzoek had gezien, was dit boekje er nooit gekomen. Heel veel dank daarvoor. Jou kwaliteit is een helikopter view hanteren, met als het moet inzoomen op essentiële details. Ik heb een uitstapje gemaakt naar de pijnpoli, en ben er enige tijd uit geweest vanwege zorgverlof. Je hebt mij hierin altijd ondersteund. Bedankt!

Judith: we hebben de afgelopen jaren intensief samengewerkt. Met name in de beginjaren moesten er heel veel formulieren en documenten geschreven worden voor centra en patiënten. Steeds weer volgde een grondige analyse met suggesties ter verbetering. Verder hebben we meermaals koffie/thee gedronken na de soms verhitte overlegmomenten op het AMC. Daarnaast heb je ook minder leuke klussen voor je rekening genomen, zoals syntaxen en databases controleren. Je hebt dit altijd zeer minutieus aangepakt, en was daarmee van grote waarde. Tot slot kan ik ons gezamenlijke "peppi en kokki" avontuur niet onvermeld laten. Middagen hebben we besteed aan het bouwen van een enorm databestand waarbij alle syntaxen moesten kloppen. Bedankt!

Collega's

Coronel Instituut AMC: bedankt voor jullie fijne collegialiteit en input voor mijn onderzoek. Merel en Inge: bedankt voor de fijne tijd als kamergenoot.

Heliomare R&D: bedankt voor de support de afgelopen jaren. Elmar: het was fijn om samen te sporten in de gym/duinen van Wijk aan Zee en mijn onderzoeks-issues met jou te bespreken. Richard: bedankt voor al je hulp bij de grafische zaken in mijn artikelen, de kaft van dit boekje, en alle andere gezellige sociale momenten.

Heliomare pijnpoli: in een lastige fase van mijn promotieonderzoek heb ik het geluk gehad te mogen werken op jullie afdeling. Ik heb hier veel geleerd en heb genoten van jullie collegialiteit, passie, en waardering voor mijn werk.

Michel: als manager Vroege Interventie ben je gedurende het gehele traject een belangrijke schakel geweest. Onze tweejaarlijkse "Vroege Interventie professional dag" uitjes (samen ook met Kurt) waren plezant. Met name het abdijhotel Rolduc zullen we niet snel vergeten, en ook Estland (Tallinn) was zeer vermakelijk. Bedankt!

Familie/vrienden/kennissen

Tante Alie en ome Dirk: van jongs af aan kom ik bij jullie over de vloer. Gedurende een lange periode heb ik bij jullie avond gegeten en ging ik daarna training geven bij de atletiek. Bedankt voor alles wat jullie voor mij en mijn ouders hebben gedaan en nog steeds doen.

Tante Teun: toen we nog op de Fluter woonde was u onze overbuurvrouw. Als kleine mannetjes maakte we dan de oversteek en mochten bij u spelen. Toen we ouder waren hebben we jarenlang op maandag de postcodeloterij show bij u gekeken. Met

een lekkere koek en chocolademelk op schoot. Dat we na al die jaren nog steeds zo goed contact hebben vind ik heel bijzonder. Bedankt voor alles.

Vriendengroep Marleen: bedankt voor de gezellige (kerst)avonden en jullie interesse in mijn onderzoek. Ome Luuk/Wijnand/Paul: bedankt voor jullie interesse in mijn promotieonderzoek. Sjors: naast Weijert ben jij mijn beste maat. Met een goed glas speciaalbier in de ene hand en droge worst/pure chocola in de andere hand had je altijd een luisterend oor en was oprecht geïnteresseerd in mijn promotieonderzoek. Meine, Sanne, Lex, Rob, Sharon, Lianne: bedankt voor jullie steun en interesse in mijn onderzoek. Cor en Verone: wat ben ik gezegend met zulke top schoonouders en opa en oma. Jullie zijn onze steun en toeverlaat. Duizend maal dank.

Jaap: jouw weerspiegeling van de *echte* wereld en dat vreemde wetenschappelijke wereldje waar je vrij weinig mee op hebt was altijd weer grappig en verhelderend. Je was wel altijd oprecht geïnteresseerd in, en blij met, de successen gedurende mijn promotietraject. Bedankt voor je steun en interesse de afgelopen jaren.

Piet: je kan helaas niet bij mijn promotie zijn, maar ik draag je altijd bij me. Op moeilijke momenten gedurende mijn traject bedacht ik me altijd dat ik heel graag de tweede uit ons gezin wilde worden met “een boekje.” Dit is altijd een grote bron van motivatie geweest. *Fino a quando mai fratello.*

Weijert: vanaf de dag dat we geboren zijn ben je mijn beste maat. We hebben zoveel mooie dingen samen beleefd. Ik hoop dit nog heel lang samen te mogen doen.

Papa en mama: ik ben ervan overtuigd dat jullie opvoeding ervoor heeft gezorgd dat ik niet voortijdig ben afgehaakt. Met hard werken en doorzettingsvermogen kan je iets moois bereiken. Bedankt voor de onbezonnen en fijne jeugd die jullie mij, Weijert, Jaap en Piet hebben gegeven, en bedankt voor alle kansen die jullie mij in het leven hebben gegeven.

Lieke en Stan: jullie zijn mijn grootste bezit. Wat hebben we samen een lol, en wat was dat vaak een fijne afleiding naast de stress die komt kijken bij een promotieonderzoek. Love you to you.

There are nine million bicycles in Beijing. That’s a fact, like the fact that I will love you till I die. Lieve Marleen, dit fragment van Katie Melua geeft mijn gevoelens voor jou goed weer. Jij bent mijn soulmate, jij bent mijn engel. Ik hou van je.

Timo

Research Institute SHARE

This thesis is published within the **Research Institute SHARE** (Science in Healthy Ageing and healthcaRE) of the University Medical Center Groningen / University of Groningen.
Further information regarding the institute and its research can be obtained from our internet site: http://www.share.umcg.nl/

More recent theses can be found in the list below.
((co-) supervisors are between brackets)

2019

Dierselhuis EF
Advances of treatment in atypical cartilaginous tumours
(prof SK Bulstra, prof AJH Suurmeijer, dr PC Jutte, dr M Stevens)

Gils A van
Developing e-health applications to promote a patient-centered approach to medically unexplained symptoms
(prof JGM Rosmalen, prof RA Schoevers)
03.07.2019 ICPE

Notenbomer A
Frequent sickness absence; a signal to take action
(prof U Bultmann, prof W van Rhenen, dr CAM Roelen)

Bishanga DR
Improving access to quality maternal and newborn care in low-resource settings: the case of Tanzania
(prof J Stekelenburg, dr YM Kim)

Tura AK
Safe motherhood: severe maternal morbidity and mortality in Eastern Ethiopia
(prof SA Scherjon, prof J Stekelenburg, dr TH van den Akker)

Vermeiden CJ
Safe motherhood: maternity waiting homes in Ethiopia to improve women's access to maternity care

(prof J Stekelenburg, dr TH van den Akker)

Schrier E
Psychological aspects in rehabilitation
(prof PU Dijkstra, prof JHB Geertzen)

Malinakova K
Spirituality and health: their associations and measurement problems
(prof SA Reijneveld, prof P Tavel, dr JP van Dijk)

Dijkhuizen A
Physical fitness and performance of daily activities in persons with intellectual disabilities and visual impairment; towards improving conditions for participation
(prof CP van der Schans, dr A Waninge, dr WP Krijnen)

Graaf MW de
The measurement and prediction of physical functioning after trauma
(prof E Heineman, dr IHF Reininga, dr KW Wendt)

Vrijen C
Happy faces and other rewards; different perspectives on a bias away from positive and toward negative information as an underlying mechanism of depression
(prof AJ Oldehinkel, prof CA Hartman, prof P de Jonge)

Moye Holz DD
Access to innovative medicines in a middle-income country; the case of Mexico and cancer medicines
(prof HV Hogerzeil, prof SA Reijneveld, dr JP van Dijk)

Woldendorp KH
Musculoskeletal pain & dysfunction in musicians
(prof MF Reneman, prof JH Arendzen, dr AM Boonstra)

Mooyaart JE
Linkages between family background, family formation and disadvantage in young adulthood
(prof AC Liefbroer, prof F Billari)

Maciel Rabello L
The influence of load on tendons and tendinopathy; studying Achilles and patellar tendons using UTC
(prof J Zwerver, prof RL Diercks, dr I van den Akker-Scheek, dr MS Brink)

Holvast F
Depression in older age
(prof PFM Verhaak, prof FG Schellevis, prof RC Oude Voshaar, dr H Burger)

For earlier theses visit our website

Extremities, Pain and Disability (EXPAND)

Wetenschappelijk onderzoek afdeling Revalidatiegeneeskunde – Centrum voor Revalidatie UMCG

EXPAND
Extremities, Pain and Disability

Missie: EXPAND draagt bij aan participatie en kwaliteit van leven van mensen met aandoeningen en amputaties van de extremiteiten of met pijn aan het bewegingsapparaat.

EXPAND omvat twee speerpunten: onderzoek naar aandoeningen aan en amputaties van extremiteiten met nadruk op stoornissen, activiteiten en participatie en onderzoek naar chronische pijn en arbeidsparticipatie. EXPAND draagt bij aan het UMCG-brede thema Healthy Ageing.

Research Department of Rehabilitation Medicine – Center for Rehabilitation UMCG

EXPAND
Extremities, Pain and Disability

Mission: EXPAND contributes to participation and quality of life of people with conditions and amputations of the extremities and musculoskeletal pain.

EXPAND focuses on two spearheads: research on the conditions and amputations of the extremities with emphasis on body functions and structures, activities and participations, and chronic pain and work participation. EXPAND contributes to Healthy Aging, the focus of the UMCG.