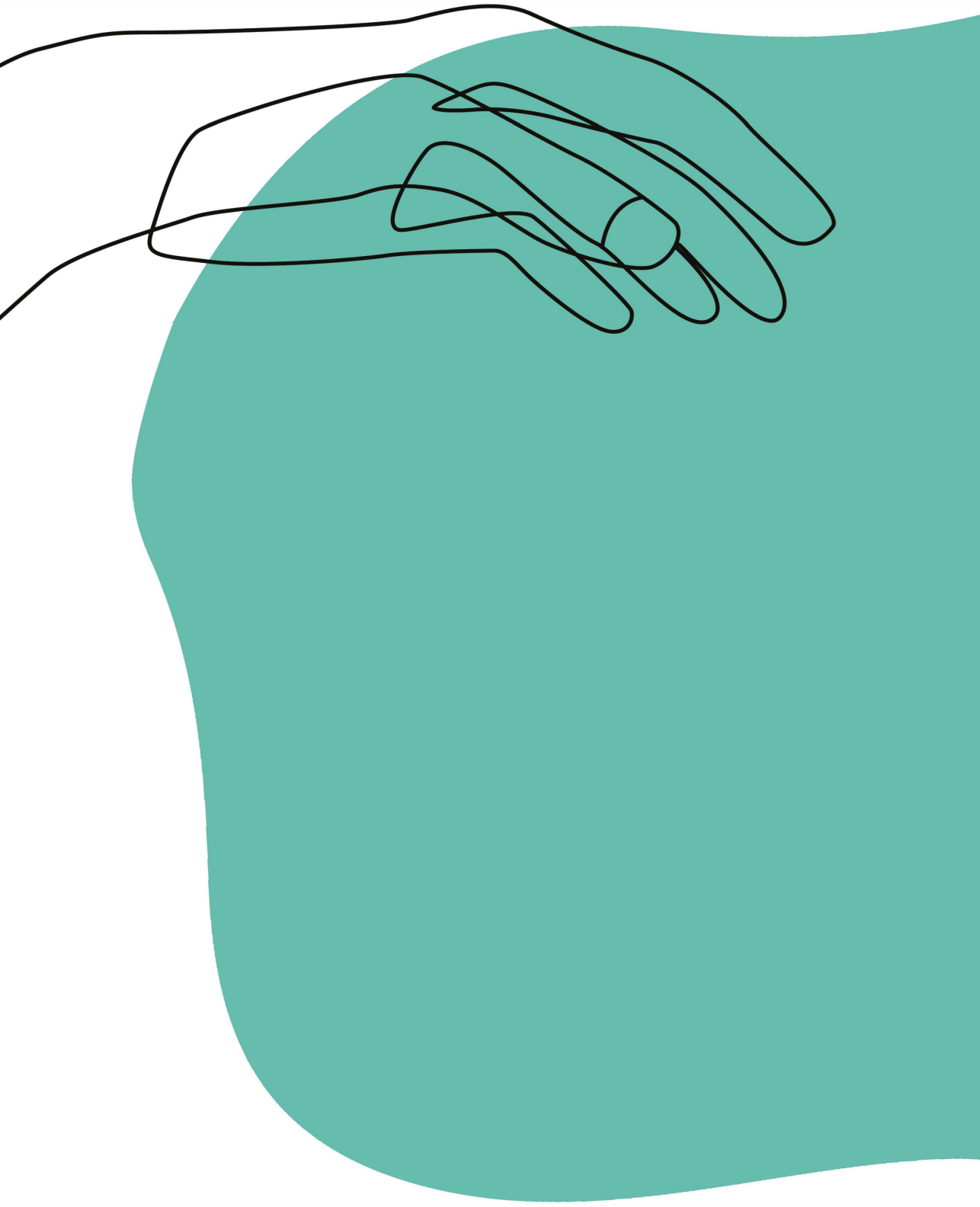
Musculoskeletal complaints in primary care

Constraining healthcare costs, rethinking the deployment of healthcare professionals

Sylvia Pellekooren



Musculoskeletal complaints in primary care

Constraining healthcare costs,
rethinking the deployment of healthcare professionals

Sylvia Pellekooren

This PhD thesis was embedded within Amsterdam Movement Sciences Research Institute, at the department of Neuro mechanics, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, the Netherlands.

Financial support was received from the Dutch Association for Manual Therapy (Nederlandse Vereniging voor Manuele Therapie) and the van Meenen middelen. The printing of this thesis was financially supported by the Scientific College Physical Therapy (WCF) of the Royal Dutch Society for Physical Therapy (KNGF).

Cover design and layout: © evelienjagtman.com
Printed by: Ridderprint | www.ridderprint.nl

ISBN: 978-94-6483-437-6
http://doi.org/10.5463/thesis.242

VRIJE UNIVERSITEIT

Musculoskeletal complaints in primary care

Constraining healthcare costs, rethinking the deployment of healthcare professionals

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor of Philosophy aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr J.J.G. Geurts,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Gedrags- en Bewegingswetenschappen
op woensdag 24 januari 2024 om 11:45 uur
in een bijeenkomst van de universiteit,
De Boelelaan 1105

door

Sylvia Pellekooren
geboren te Boskoop

promotoren:	prof.dr. A.L. Pool-Goudzwaard
	prof.dr. R.W.J.G. Ostelo
copromotor:	prof.dr. M.W. van Tulder
commissieleden:	dr. G.G.M. Scholten-Peters
	prof.dr. H. Schers
	prof.dr. C. Veenhof
	prof.dr. P. van der Wees
	dr. F. van Nassau

Contents


1

General introduction

General introduction

Challenges in healthcare

Worldwide policy makers are challenged to account for rising healthcare costs and increased healthcare demand. Also in the Netherlands there is a growing concern how to maintain high-quality and accessible care while keeping costs in check.[1] Access to care is under pressure as the demand for care is rising fast, due to an aging population and an increasing number of chronically ill people. Not only at the policy level, but also in clinical practice challenges exist. The workload in the healthcare sector is high, causing healthcare workers to leave the sector, which further increases labor market shortages and hence workloads.[1] Simply investing more money in the care sector and employing more healthcare workers is not a solution, as 13% of the national income is spend on care and welfare -while public funds are also needed for other social ends, such as education-, and one in six workers in the Netherlands has a job in the healthcare sector.[1] Now is the time to constrain healthcare cost and to rethink the way we deploy healthcare workers within the complex landscape of the Dutch healthcare system.

Dutch healthcare system

The Dutch healthcare system is characterized by a demand-driven system with regulated competition and elements of both public and private insurance. All Dutch residents are mandated to have a basic health insurance package, which includes - amongst others - a standard package of care provided by GPs, with GP consultations being fully reimbursed.[2] The vast majority of all healthcare pathways start with consulting a GP, as in the Dutch healthcare system, patients are registered at one GP only, who is the gatekeeper for patients' referral to primary and secondary healthcare facilities.[3] Patients with musculoskeletal complaints, however, can also consult a physiotherapist directly through Direct Access Physiotherapy. In 2018, 68% of all patients consulted a physiotherapist without a referral, and of those with a referral, 72% were referred by a GP.[4]

Musculoskeletal complaints in General Practice

Of all contacts within general practice, 14.6% relate to musculoskeletal conditions.[5] Besides representing a large patient population within GP practice, patients with musculoskeletal conditions are also known to contribute to high healthcare costs. In 2017, the costs of musculoskeletal conditions in the Netherlands were estimated to be 6,3 billion euros, equaling about 7.5% of the total Dutch health care budget[6] and total costs for musculoskeletal complaints in general practices amounts 342,5 million euros, which is 8.8% of total cost made in general practice.[7] Estimates based

on demographics (i.e., population growth and ageing) and changing healthcare use (i.e., medical technology and increased welfare) show that healthcare costs for musculoskeletal conditions are expected to increase by an average of 2.85 per cent per year.[6]

Healthcare utilization and cost

International publications report that a predominant factor in the increase in healthcare costs in general practice appears to be high healthcare utilization (i.e., GP visits, physiotherapy visits, and outpatient visits), rather than initial unit costs, such as cost per visit.[8] Moreover, a considerable part of resource utilization seems to be attributable to a small number of patients, so-called high-cost users[9,10]; approximately 35% of healthcare cost were concentrated among the 4% persistent high-cost utilizers.[9] Gaining insight into which patients in Dutch general practice are so-called 'high cost-users' would be useful in guideline development and policymaking, thereby possibly reducing healthcare utilization and healthcare costs. **Chapter 2** evaluates healthcare utilization and cost of GP-guided care in patients with musculoskeletal complaints at a disaggregated level and identifies predictive factors for having higher healthcare cost. Healthcare costs of GP led care pathways in patients with musculoskeletal complaints were described using healthcare utilization data derived from the Pharmo Institute database, which contains longitudinal and real-world individual patient data and included GPs electronic medical records of 2,1 million patients. In addition, these data were used to identify a set of possible predictors of having higher healthcare cost among patients with musculoskeletal complaints in Dutch general practice by means of a linear regression analysis.

Healthcare utilisation is also dependent on to what extend patients are satisfied with treatment received, as satisfied patients seem to gain more benefit from healthcare,[11] keep their appointments,[12,13] and are more adherent to treatment.[14,15] Satisfaction outcomes also play an important role in health policy, which influences healthcare providers' interaction with patients and thereby healthcare costs.[16,17] Although patient satisfaction seems an important parameter and is often measured in the assessment of healthcare interventions and in new models of care,[18,19] the best method of measuring patient satisfaction is unclear.[20,21] This is partly due to problems with the content of measurement instruments used, which do not always adequately reflect the construct being measured. **Chapter 3** assesses the content validity of Patient-Reported Outcome Measures that aim to measure patient satisfaction in primary care for musculoskeletal complaints by means of a systematic review using the COSMIN methodology.

Advanced practice physiotherapy within Dutch primary care

An ageing population and an increasing number of chronically ill patients does not only lead to more healthcare utilisation within general practices, but also to more complex healthcare demands.[22-24] With more complex and demanding patients, taking over tasks from secondary care facilities, and more administrative tasks GPs workload is increasing, while at the same time the number of colleagues is decreasing.[23] Ultimately, this high workload leads to suboptimal diagnostics and more referrals to diagnostic imaging and secondary care facilities, and thus higher healthcare costs.[24] To cope with this, GPs organise support and expertise within their practices by collaborating with other healthcare providers, such as nurse practitioners in the care of patients with chronic conditions or geriatrics specialist in caring for vulnerable elderly.[25] Following these examples, several initiatives have been taken to relieve GPs in the care for patients with musculoskeletal conditions. One such initiative is the deployment of Advanced Practitioner Physiotherapists (APPs), also referred to as Extended Scope Practitioners (ESPs),[26,27] who take over tasks traditionally performed by GPs, such as setting and communicating diagnoses, setting care pathways, and ordering diagnostic imaging and laboratory tests.

APP was introduced in the Netherlands in 2017 and by 2021, 650 APPs were trained, some of whom are already active within general practice and other healthcare settings throughout the Netherlands.[28] Since APP is a relatively new healthcare provider it is unclear how APP will find its footing in the existing healthcare landscape. Several qualitative studies[29–33] have explored the barriers and facilitators of the implementation of APP model of care within secondary healthcare settings, concluding that knowledge, skills, availability of APP, motivation, and experience all have a large impact upon its successful implementation. Although these studies are undoubtedly of great value in terms of improving the implementation of APP, their outcomes are not necessarily applicable to the Dutch primary care setting. This is because these studies focused on the implementation of APP within secondary care facilities in other healthcare systems, and, as such, those barriers and facilitators that are specific to the Dutch primary care setting may not have been addressed. **Chapter 4** evaluates the introduction of an APP model of care in Dutch general practice by conducting an explorative and interpretive qualitative study among Dutch APPs and GPs who were in various stages of implementing a musculoskeletal APP care model within a primary care setting.

In other counties, where APP is deployed for a longer period, the body of evidence is growing on health effects, diagnostic accurateness, and healthcare cost. Recently published reviews report that the use of APPs contributes to accessibility of care with similar health effects and healthcare costs.[26,34] Despite positive results in the international literature, little is known about the effects of the deployment of APP in Dutch primary care.

Chapter 5 explores the deployment of APPs in general practice by describing APP patient population, evaluating health effects, identifying APP led pathways and calculating associated healthcare costs, hereby providing a first impression of the deployment of APP in primary care in the Netherlands.

EQ-5D based utility values

The exploratory study on the deployment of APP included a set of outcome measures including, amongst others, health related quality of life, which was assessed using EQ-5D.[35] This preference-based quality-of-life measure is used to estimate utility values that represent the preferences of the general population of a country for given health states.[36] These utility values, which combine both the quality and quantity of life into a single outcome,[37] are needed for estimating Quality-Adjusted Life-Years (QALYs) in cost effectiveness analysis. Unfortunately, EQ-5D data are not always available in clinical trials,[38] as higher priority is sometimes given to condition-specific measures that assess more clinically relevant outcomes.[39] This issue is even more pronounced in real-world data (e.g., electronic health records such as used in **Chapter 1**), as these data are typically collected for clinical purposes only.[40]

In the absence of the EQ-5D or another generic preference-based quality-of-life measure, a condition-specific measure might be used to predict utility values.[41] **Chapter 6** evaluates the predictive performance of different types of prediction modeling to explore which model performs best in predicting EQ-5D based utility values by using the Oswestry Disability Index, and reports the results of Ordinary Least Squares regression and Tobit regression (i.e., censored, or truncated regression) models in predicting EQ-5D-3L utility values using the Oswestry Disability index for use in cost-effectiveness analyses.

Although, predicted EQ-5D-based utility values can be validly used in cost-effectiveness analyses, there is no consensus yet whether regression modelling is the best way to do so.[39] Evidence suggests that response mapping approaches perform better than regression models and might be better at preventing regression to the mean,[42] because they aim to align the scales between instruments so that the distributions of their responses are matched.[39,43] Hence, response mapping approaches might result in more accurate estimates of individual scores on the target instrument. **Chapter 7** explores if cost-utility results are valid when mapping Oswestry Disability Index responses to EQ-5D utility values and outperform the developed regression models..

Objectives and outline of this thesis

In sum this thesis this thesis includes three topics with the following objectives:

Part I. Health care utilization and cost

- Chapter 2 evaluates healthcare utilization and cost of GP-guided care in patients with musculoskeletal complaints at a disaggregated level and identifies a set of predictive factors for having higher healthcare cost.
- Chapter 3 assesses the content validity of measurement instruments that measure patient satisfaction, as this is a frequently measured parameter of quality of care that influences healthcare utilization.

Part II. Advanced practice physiotherapy within Dutch primary care

- Chapter 4 evaluates the introduction of an APP model of care in Dutch general practice.
- Chapter 5 explores the deployment of APPs in general practice by identifying APP patient population and evaluating APP-led health care pathways and associated cost.

Part III. EQ-5D based utility values

- Chapter 6 evaluates the predictive performance of different types of prediction modeling (i.e., censored, or truncated regression) to explore which method performs best in predicting EQ-5D based utility values by using the Oswestry Disability Index.
- Chapter 7 explores if cost-utility results are valid when mapping Oswestry Disability Index responses to EQ-5D utility values and outperform the developed regression models

References

1. Ministry of Public Health Welfare and Sport. Integraal Zorgakkoord: "Samen Werken Aan Gezonde Zorg."; 2022.
2. Ministry of Public Health, Welfare and Sport. Accessed March 27, 2022. https://english.zorginstituutnederland.nl/about-us/publications/publications/2016/01/31/healthcare-in-the-netherlands
3. Dutch College of General Practitioners. Accessed March 25, 2022. https://www.nhg.org/english/gp-dutch-healthcare-system
4. Veldkamp R, Meijer W. Zorg Door de Fysiotherapeut. Nivel Zorgregistraties Eerste Lijn, Jaarcijfers 2021 En Trendcijfers 2018-2021.; 2022. Accessed March 22, 2023. https://www.nivel.nl/sites/default/files/bestanden/1004328.pdf
5. Flinterman L, Groenewegen P, Verheij R. Zorglandschap En Zorggebruik in Een Veranderende Eerste Lijn.; 2018.
6. Ministry of Health Welfare and Sport. Public Health Foresight Study 2018. https://www.vtv2018.nl/zorguitgaven.
7. National Institute for Public Health and the Environment. Kosten van ziekten. https://www.vzinfo.nl/kosten-van-ziekten.
8. Burgess R, Hall J, Bishop A, Lewis M, Hill J. Costing Methodology and Key Drivers of Health Care Costs Within Economic Analyses in Musculoskeletal Community and Primary Care Services: A Systematic Review of the Literature. J Prim Care Community Health. 2020;11:215013271989976. doi:10.1177/2150132719899763
9. Lentz TA, Harman JS, Marlow NM, Beneciuk JM, Fillingim RB, George SZ. Factors associated with persistently high-cost health care utilization for musculoskeletal pain. PLoS One. 2019;14(11):e0225125. doi:10.1371/journal.pone.0225125
10. Wammes JJG, van der Wees PJ, Tanke MAC, Westert GP, Jeurissen PPT. Systematic review of high-cost patients' characteristics and healthcare utilisation. BMJ Open. 2018;8(9):e023113. doi:10.1136/bmjopen-2018-023113
11. Fitzpatrick R. The assessment of patient satisfaction. In: Jenkinson J, ed. Assessment and Evaluation of Health and Medical Care. Open University Press; 1997:85-102.
12. Bigby J. Medical Consequences of Missed Appointments. Arch Intern Med. 1984;144(6):1163. doi:10.1001/archinte.1984.00350180075012
13. Frankel S, Farrow A, West R. Non-attendance or non-invitation? A case-control study of failed outpatient appointments. BMJ. 1989;298(6684):1343-1345. doi:10.1136/bmj.298.6684.1343
14. Kincey J, Bradshaw P, Ley P. Patients' satisfaction and reported acceptance of advice in general practice. J R Coll Gen Pract. 1975;25:558-566.
15. Willson P, McNamara JR. How perceptions of a simulated physician-patient interaction influence intended satisfaction and compliance. Soc Sci Med. 1982;16(19):1699-1704. doi:10.1016/0277-9536(82)90095-8
16. Bachman J. The Problem with Patient Satisfaction Scores. Fam Pract Manag. 2016;23:23-27.
17. Kupfer JM, Bond EU. Patient Satisfaction and Patient-Centered Care. JAMA. 2012;308(2):139. doi:10.1001/jama.2012.7381
18. Desmeules F, Roy JS, MacDermid JC, Champagne F, Hinse O, Woodhouse LJ. Advanced practice physiotherapy in patients with musculoskeletal disorders: a systematic review. BMC Musculoskelet Disord. 2012;13(1):107. doi:10.1186/1471-2474-13-107

19. Fennelly O, Blake C, Desmeules F, Stokes D, Cunningham C. Patient-reported outcome measures in advanced musculoskeletal physiotherapy practice: a systematic review. Musculoskeletal Care. 2018;16(1):188-208. doi:10.1002/msc.1200
20. Batbaatar E, Dorjdagva J, Luvsannyam A, Amenta P. Conceptualisation of patient satisfaction: a systematic narrative literature review. Perspect Public Health. 2015;135(5):243-250. doi:10.1177/1757913915594196
21. Hush JM, Cameron K, Mackey M. Patient Satisfaction With Musculoskeletal Physical Therapy Care: A Systematic Review. Phys Ther. 2011;91(1):25-36. doi:10.2522/ptj.20100061
22. Schäfer W, van den Berg M, Groenewegen P. De werkbelasting van huisartsen in internationaal perspectief. Huisarts Wet. 2016;59:94-101.
23. van der Horst H, Bindels P, Assendelft P, et al. Hoogste tijd voor minder patiënten per huisarts. Huisarts Wet. 2018;61(3):30-31. doi:10.1007/s12445-018-0056-9
24. National General Pactitioners Association. Meer Tijd Voor de Patient.; 2018.
25. van Dongen J, van Bokhoven L, Goossens W, Daniëls R, van der Weijden T, Beurskens S. Steeds complexere zorg vraagt om teamwerk. Huisarts Wet. 2019;62(11):60-63. doi:10.1007/s12445-019-0305-6
26. Vedanayagam M, Buzak M, Reid D, Saywell N. Advanced practice physiotherapists are effective in the management of musculoskeletal disorders: a systematic review of systematic reviews. Physiotherapy (United Kingdom). 2021;113:116-130. doi:10.1016/j.physio.2021.08.005
27. Tawiah AK, Desmeules F, Finucane L, et al. Advanced practice in physiotherapy: a global survey. Physiotherapy (United Kingdom). 2021;113:168-176. doi:10.1016/j.physio.2021.01.001
28. Eskens M, Hallegraeff H. Beroepscompetentieprofiel Extended Scope Specialisten.; 2022.
29. Jones A, Powell T, Watkins D, Kelly D. Realising their potential? Exploring interprofessional perceptions and potential of the advanced practitioner role: a qualitative analysis. BMJ Open. 2015;5(12):e009740. doi:10.1136/bmjopen-2015-009740
30. Morris J, Grimmer K, Ashman B, et al. Principles to guide sustainable implementation of extended-scope-of-practice physiotherapy workforce redesign initiatives in Australia: stakeholder perspectives, barriers, supports, and incentives. J Multidiscip Healthc. Published online June 2014:249. doi:10.2147/JMDH.S59473
31. Dawson LJ, Ghazi F. The experience of physiotherapy extended scope practitioners in orthopaedic outpatient clinics. Physiotherapy. 2004;90(4):210-216. doi:10.1016/j.physio.2004.06.001
32. Shaw BR, Heywood SE, Page CJ, et al. Advanced musculoskeletal physiotherapy: Barriers and enablers to multi-site implementation. Musculoskeletal Care. 2018;16(4):440-449. doi:10.1002/msc.1358
33. Tawiah AK, Borthwick A, Woodhouse L. Advanced Physiotherapy Practice: A qualitative study on the potential challenges and barriers to implementation in Ghana. Physiother Theory Pract. 2020;36(2):307-315. doi:10.1080/09593985.2018.1484535
34. Lafrance S, Demont A, Thavorn K, Fernandes J, Santaguida C, Desmeules F. Economic evaluation of advanced practice physiotherapy models of care: a systematic review with meta-analyses. BMC Health Serv Res. 2021;21(1). doi:10.1186/s12913-021-07221-6
35. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy (New York). 1990;16(3):199-208. doi:10.1016/0168-8510(90)90421-9
36. Froberg DG, Kane RL. Methodology for measuring health-state preferences–I: Measurement strategies. J Clin Epidemiol. 1989;42(4):345-354. doi:10.1016/0895-4356(89)90039-5
37. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. Measuring and Valuing Health Benefits for Economic Evaluation. Vol 1. Oxford University Press; 2016. doi:10.1093/med/9780198725923.001.0001

38. Gianola S, Frigerio P, Agostini M, et al. Completeness of Outcomes Description Reported in Low Back Pain Rehabilitation Interventions: A Survey of 185 Randomized Trials. Physiotherapy Canada. 2016;68(3):267-274. doi:10.3138/ptc.2015-30
39. Fayers PM, Hays RD. Should Linking Replace Regression When Mapping from Profile-Based Measures to Preference-Based Measures? Value in Health. 2014;17(2):261-265. doi:10.1016/j.jval.2013.12.002
40. Fatoye F, Gebrye T, Odeyemi I. Real-world incidence and prevalence of low back pain using routinely collected data. Rheumatol Int. 2019;39(4):619-626. doi:10.1007/s00296-019-04273-0
41. Longworth L, Rowen D. Mapping to obtain EQ-5D utility values for use in nice health technology assessments. Value in Health. 2013;16(1):202-210. doi:10.1016/j.jval.2012.10.010
42. Thompson NR, Lapin BR, Katzan IL. Mapping PROMIS Global Health Items to EuroQol (EQ-5D) Utility Scores Using Linear and Equipercentile Equating. Pharmacoeconomics. 2017;35(11):1167-1176. doi:10.1007/s40273-017-0541-1
43. Dorans NJ. Linking scores from multiple health outcome instruments. Quality of Life Research. 2007;16(S1):85-94. doi:10.1007/s11136-006-9155-3

PART I

Healthcare utilization and costs of musculoskeletal complaints


2

Predicting Direct Healthcare Costs of General Practitioner Guided Care in Patients with Musculoskeletal Complaints

Pain. 2023; doi:10.1097/j.pain.0000000000003028

Pellekooren S, Ben ÂJ, van Dongen JM, Pool-Goudzwaard AL, van Tulder MW, Van den Berg JM, Ostelo RWJG.

Abstract

Background

Information on healthcare utilization and cost of general practitioner (GP)-guided care in patients with musculoskeletal complaints is important for keeping healthcare affordable and accessible. Previous studies provided insight into the resource use and costs of patients with musculoskeletal conditions in a broad range of settings, as well as predictive factors of having high(er) costs. However, previous studies assessing the costs of musculoskeletal conditions in general practices either only presented costs at the aggregate level and not per musculoskeletal condition separately or reported on low-back pain patients only.

Objectives

To describe healthcare utilization and costs of GP-guided care for patients with various musculoskeletal complaints separately, and to predict higher direct healthcare costs in GP guided care for patients with musculoskeletal complaints.

Methods

A registry-based study was performed using data from electronic medical records of 2,118,603 adult patients treated in general practice, which is 13% of all adult patients registered with a GP practice in the Netherlands. Healthcare costs of GP-guided care included all healthcare resources used by patients due to a musculoskeletal condition in 2018. Data were extracted from the database with a 1-year follow-up and descriptively analyzed. A General Linear Model was developed to identify a set of predictive factors for having higher healthcare costs.

Results

In total, 403,719 patients were included, of whom 92% only received a single consultation. The number of referrals varied widely across the different types of complaints. Total annual healthcare costs amounted to €39,180,531, of which the key cost driver was referrals. Referrals to primary care providers accounted for the largest part of referral-related costs. For all musculoskeletal conditions combined, mean annual healthcare cost per patient were €97 (SEM=€0.18). High age, being female, low social economic status, spine complaints, high number of musculoskeletal diagnoses, and a high comorbidity score were predictive of having higher healthcare costs and explained 0.7% of the variance.

Conclusions

This study showed that mean annual healthcare cost of GP-guided care in patients with musculoskeletal conditions were relatively low and did not differ considerably across conditions. Investigated predictive factors explained a negligible part of the variance in cost. Thus, it is unclear which factors do explain high healthcare costs in this population.

Introduction

Most patients with musculoskeletal conditions are treated in a primary care setting. Physiotherapists and general practitioners (GP) are often the first point of contact. In the Netherlands, 14.6% of all contacts with a GP are related to musculoskeletal conditions.[12] Besides representing a large group of patients in primary care, patients with musculoskeletal conditions also contribute to high healthcare costs. In 2017, the total healthcare costs of musculoskeletal conditions in the Netherlands - including primary, secondary, and tertiary care - were estimated at 6.3 billion euros, equaling about 7.5% of the total Dutch healthcare budget.[29] According to a report by the Dutch Ministry of Health, Welfare and Sports, the total costs of musculoskeletal complaints in general practices amount to 342.5 million euros, which is 8.8% of the total GP costs.[29] Estimations based on demographic development (i.e., population growth and aging) and changing healthcare use (i.e., medical technology and increased welfare) forecast that the healthcare costs of musculoskeletal conditions will increase during the upcoming decades, with an average of 2.85% per year.[30] Research also indicates that a considerable part of these costs is attributable to a relatively small group of patients, so-called high-cost users.[18, 36] Previous studies provided insight into the resource use and costs of patients with musculoskeletal conditions in a broad range of settings, as well as predictive factors of having high(er) costs.[3,18,36] However, previous studies assessing the costs of musculoskeletal conditions in general practices either only presented costs at the aggregate level and not per musculoskeletal condition separately or reported on low-back pain patients only. The aim of this study is to describe healthcare utilization and costs of GP-guided care for patients with various musculoskeletal complaints separately, and to predict higher direct healthcare costs in GP guided care for patients with musculoskeletal complaints.

Methods

Setting and population

The Dutch healthcare system is characterized by a demand-driven system with regulated competition and elements of both public and private insurance. All Dutch residents are mandated to have a basic health insurance package, which includes - amongst others - a standard package of care provided by GPs, with GP consultations being fully reimbursed.[21] The vast majority of all healthcare pathways start with consulting a GP, as in the Dutch healthcare system, patients are registered at one GP only, who is the gatekeeper for patients' referral to primary and secondary healthcare facilities.[7] Patients with musculoskeletal complaints, however, can also consult a physiotherapist

directly through Direct Access Physiotherapy. In 2018, 68% of patients consulted a physiotherapist without a GP referral.[35] In describing healthcare utilization and costs for patients with musculoskeletal complaints, consultations related to Direct Access Physiotherapy were not included as this care is guided by physiotherapist and not by GPs. Other allied healthcare professionals are only accessible via a GP referral. One full-time GP provides medical care for approximately 2,350 patients.[7]

Source of data

Data was used from the PHARMO Database Network.[27] This database contains longitudinal real-world patient data that is retrieved, amongst others, from GPs' electronic patient files, pharmacies, and laboratories throughout the Netherlands, and includes electronic medical records of 2,118,603 Dutch adult patients, which is 13% of all adult patients registered with a GP practice in the Netherlands.[11] Data were recorded as part of routine clinical practice at GP offices and included prospectively collected anonymous patient-level data on consultations, medication, and referrals by GPs. This information was extracted from problem lists, and journal text. Patients of 18 years and older were included in this study when an International Classification of Primary Care (ICPC) code[17] that represents a musculoskeletal complaint was assigned in 2018. Follow-up data had to be available for a minimum duration of one year after allocating the ICPC code (i.e., index date). An overview of included ICPC codes is presented in Appendix I. No ethical approval was needed, as this study analyzed already existing completely anonymous data.

Healthcare utilization

Healthcare utilization of GP-guided care was defined as all healthcare resources used by patients due to a musculoskeletal condition guided by a GP, and included: GP care itself (office-based consultations, home visits, consultations by phone/email, and consultations by a GP-based nurse specialist or physician assistant), referrals to other healthcare providers (to primary care -i.e., physiotherapy, manual therapy, occupational therapy, exercise therapy, dietetics, mental health services, and orthopedic technician-, referrals to secondary care -i.e., consultation medical specialist-, and referrals to diagnostic imaging i.e., echography, X-ray, CT, MRI), and prescribed medication. Referrals to primary care involved treatment episodes with primary care givers (e.g., physical therapist), while referrals to secondary care only involved a single consultation by a secondary care provider (e.g., medical specialist). Secondary care referrals did not include any follow-up appointments, examinations, surgery, possible admissions, or cost-related to the provision of specialized treatment, as this care is guided by medical specialists and not by GPs and was therefore not included in this dataset. Although the patients' actual attendance to referral visits could not be confirmed it was assumed that

all patients attended their referral appointment, as secondary care in the Netherlands is only accessible through a referral by the GP, is fully reimbursed by the basic health care insurance which is mandatory for all Dutch residents, and no-show rates in the Netherlands are relatively low. Referrals that were not related to a musculoskeletal complaint (e.g., an ophthalmologist) were excluded from further analysis. Prescribed medication was classified according to the Anatomical Therapeutic Chemical (ATC) classification system.[37]

Direct healthcare costs

Direct healthcare costs were defined as costs of the healthcare utilization described above. GP care was valued using Dutch standard prices derived from the Dutch guidelines for costing studies.[13,22] For referrals to primary care providers, the patients' number of contacts could not be retrieved from the database and were estimated based on national estimates of the average number of contacts for a broad range of care providers (e.g., physiotherapy and exercise therapy).[11, 24] For social work and primary care-based psychologists, the number of contacts was based on data from the Netherlands Institute of Mental Health and Addiction.[33] Dutch standard prices were used to value healthcare utilization.[13,22] Medication use was valued using prices published by the National Health Care Institute.[23] All costs were expressed in Euros 2021. The time horizon for costs was 1 year. See Table 1 for more details**.**

Predictive factors

Based on previous literature, a set of potential predictive factors were used to predict higher direct healthcare costs [18,37], but only predictive factors that were registered in the database could be included. Factors included were age (years), sex (male/female), region of complaint (i.e., lower extremity, upper extremity, and spine), type of complaint (i.e., based on guidelines for general practitioners [8]), the Chronic Disease Score (i.e., overall score retrieved by an algorithm for co-morbidities which is based on assigned ICPC codes and prescribed medication), cardiovascular disease (yes/no), diabetes mellitus (yes/no), smoking (yes/no), obesity (yes/no), the total number of musculoskeletal diagnoses, depression (yes/no), and neighborhood Social Economic Status (SES) (i.e., low, middle, or high).[15,33] Some of the variables included were generated based on other ICPC codes assigned to the patients. To illustrate, cardiovascular disease (yes/no), diabetes mellitus (yes/no), smoking (yes/no), and obesity (yes/no) were based on whether patients received ICPC codes for these conditions and were considered from the time of inception of the database (i.e., 2008) up until the index date. For depression, a similar approach was used, but only the 6-month period before the index date was considered. The chronic disease score is defined as a set of scoring rules that render a score between 0 and

41, reflecting chronic disease severity based on a single drug or combinations of drugs. The CDS is associated with subsequent-year hospitalization and mortality: a CDS of 7 or greater is associated with a 5-fold increase in the risk of hospitalization and a 10-fold increase in the risk of dying.[6,26] Further details on variables region of complaint and type of complaint are described in Appendix I. Further details on the variables SES, cardiovascular disease, diabetes mellitus, smoking, and obesity are shown in Appendix II. The total number of musculoskeletal complaints was based on the number of musculoskeletal-related ICPC codes that patients received during the 6-month period before the index date.

Statistical Analysis

Seven percent of all patients had one or more missing values. These were imputed using the Multivariate Imputation by Chained Equations package (MICE).[4,37] The MICE framework uses Predictive Mean Matching (PMM) to generate several datasets with complete data randomly sampled from the observed data after matching observations with missing values with those with complete observed data (referred to as donor observations). PMM assumes that the distribution of the missing data is the same as the observed data and avoids imputing implausible values (negative costs).[4,37] Five complete datasets were generated using an imputation model that included ICPC-code, sex, age, SES, the total number of musculoskeletal complaints, Chronic Disease Score, cardiovascular disease, depression, obesity, smoking, type of referral, type of medication, type of GP consultation, and observation linked to the episode of interest (i.e., the GP consultation, referral, or prescribed medication is related to the musculoskeletal compliant) (yes/no). After data imputation, healthcare utilization and costs of GP-guided care were descriptively analyzed using means and SEMs (continuous variables) and counts and percentages (dichotomous and categorical variables).

For the descriptive analyses, healthcare utilization and costs were presented for the entire study population, as well as for patients who belonged to one of the four cost quartiles and those who belonged to the top 5% of high-cost users. Results were presented for all musculoskeletal conditions combined, by region of complaint (i.e., spine, lower extremity, upper extremity) and for the 5 most common types of complaint based on Dutch GP guidelines (i.e., hand/wrist complaints, knee complaints, low-back pain, low-back pain with radiation, and shoulder complaints).[8] More detailed information on ICPC codes included per region and per type of complaint are presented in Appendix I.

The dependent variable in the regression analysis was having higher direct healthcare costs, which was included as a continuous outcome. To predict higher direct healthcare costs of GP-guided care, each of the five complete datasets was analyzed separately as outlined below, after which results were pooled using Rubin's rules.[4,20] Total direct healthcare costs were regressed upon the possible predictive factors using a General Linear Model (GLM) with a Gamma distribution and an identity link. A GLM was chosen, because of the skewed distribution of healthcare costs.[1] Prior to the analyses, assumptions of a GLM with a Gamma distribution were checked (i.e., normally distributed residual variance, homoscedasticity, influential cases, and outliers). The model was trained and evaluated using the Caret package.[16] K-fold cross-validation was used to internally validate the model.[2] Predictors were selected using a bi-directional stepwise selection procedure,[31] using the Akaike Information Criterion (i.e., the trade-off between the goodness of fit of the model and the simplicity of the model),[5] with a 5% significance level. Stepwise selection combines the elements of forward and backward selection by sequentially adding variables, based on the most contributing predictors, and omitting variables that no longer provide an improvement in the model fit after adding a new variable to the model. The final model only included case-mix variables that increased the predictive value. The overall performance of the model was assessed using the RMSE (i.e., the absolute fit of the model) and the adjusted R2 (i.e., the relative fit of the model). Analyses were conducted in R (version 4.0.0).

Table 1. Costing details consultations, referrals, and medication

	Unit cost (€)	Number of units	Total cost (€)
Consultations[1]			
Office based consult	36.04	1	36.04
Home visit	54.60	1	54.60
Consultation by phone/email	18.56	1	18.56
Consultation by general-practice-based nurse specialists	18.56	1	18.56
Primary care referrals[1]			
Physiotherapy	36.04	7,5	270.3
Occupational therapy	36.04	8,1	219.92
Manual therapy	36.04	7,5	270.3
Exercise therapy	37.13	10,5	389.87
Podotherapy	279.23	1	279.23
Medical devices	1050	1	1050
Dietetics	32.41	3.2	103.71
Psychologist	100.81	8	806.48
Social worker	70.98	8	567.84
Secondary care referrals			
Consultation medical specialist	132.68	1	132.68
Imaging refarrals			
Radiological	170,95	1	170.95
Echography	93,60	1	93.60
Medication[2]			
PCM	0.07	1	0.07
NSAIDS	0.08	1	0.08
Opioids weak	0.06	1	0.06
Opioids strong	2.04	1	2.04
Co analgesic	0.05	1	0.05
Injections*	4.77	1	4.77

Costs are presented in euros 2021.

[1] Unit cost are based on Dutch guidelines for costing studies [13,22]

[2] Unit cost are based on prices published by the National Health Care Institute [23]

*The GP consultation during which the injection was administered is counted separately under consultation.

Results

In total, 403,719 patients were included. The mean age was 53 (SD=18) and 227,101 (56%) patients were female. The most common region of complaints was the lower extremities (28%), and the most common type of complaint was shoulder complaints 34,579 (9%). See Table 2 for more characteristics and Appendix III and IV for more details for region and type of complaint.

Healthcare utilization

The most common type of GP-guided care was office-based consultations (93%). Of all patients, 92% received a single consultation only, and 21% received one or more referrals (of which 41% were to primary caregivers, 43% to a medical specialist, and 7% to diagnostic imaging). Of all referrals to a primary care provider, 23% were physiotherapy referrals, 3% were podiatry referrals and 1% were exercise therapy referrals. Medication was prescribed to 17% of patients, with NSAIDs being most prescribed (41%). The number of referrals varied widely across the different types of complaints. For example, of all patients who consulted the GP due to low back pain, 20% received a referral, of which 27% were to a secondary care facility. In case of hand/wrist complaints, 22% of patients received a referral, of which 62% made was to a secondary care facility. See Table 3 for more details for region and type of complaint.

Direct healthcare costs

In the complete study sample, the one-year mean total direct healthcare cost of GP-guided care per patient was €97 (SEM=€0.18). The one-year mean cost per patient for GP consultations and referrals (i.e., cost generated by the referral) were €43 (SEM=€0.03) and €53 (SEM=€0.18), respectively. Primary care referrals were associated with the highest cost average per patient (€33; SEM=€0.14), followed by secondary care referrals (€17; SEM=€0.07), and diagnostic imaging (€4; SEM=€0.03). For prescribed medication, the one-year mean cost per patient was €1 (SEM=€0.02). Among patients who were referred, one-year average primary care and secondary care costs were €303 (SEM=€0.51) and €155 (SEM=€0.30). Among patients referred to diagnostic imaging, the one-year average imaging cost was €177 (SEM=€0.36). More detailed information on one-year mean healthcare cost of GP-guided care per patient is shown in Table 4.

Total one-year mean costs were similar among the different regions and types of complaints except for low back pain with radiation, of which the one-year average cost per patient was €137 (SEM=€1.58) per patient. The one-year mean total direct

healthcare cost for high-cost users (i.e., top 5%) was €466 (SEM €1.12), with referrals to primary caregivers being the largest cost driver. More detailed information on costs per region and type of complaint are shown in Appendix V.

In the complete study sample (n=403,719), the one-year total direct healthcare costs of GP-guided care amounted to €39,180,531. High-cost users were responsible for 24% of these costs (i.e., €9,406,681). More detailed information one-year total direct healthcare costs of GP-guided care per region and per type of complaint is shown Table 5.

Predictive factors

Predictive factors for having higher direct healthcare costs were high age, being female, low social economic status, spine complaints, a high number of musculoskeletal diagnoses, and a high comorbidity score. The model explained 0.7% (R^2) of the variation in the outcome (i.e., higher direct healthcare costs) and the absolute fit (RMSE) was 127.8. More details on the regression model are presented in Appendix VI.

Table 2. Baseline characteristic overall musculoskeletal complaints

	Overall (n=403,719)
Sex: Female (Yes; n, %)	227,101 (56.3)
Age (mean, SD)	52.54 (18.31)
Comorbidities	
Chronic Disease Score (mean, SD)	2.78 (3.71)
Number of musculoskeletal diagnoses (mean, SD)	0.08 (0.31)
Depression (Yes; n, %)	2,225 (0.6)
Obese (Yes; n, %)	1314 (0.3)
Smoking (Yes; n, %)	1726 (0.4)
Cardiovascular (Yes; n, %)	112,979 (28.0)
Diabetes (Yes; n, %)	22758 (5.6)
Social Economic Status	
Low (n, %)	146,820 (36.4)
Middle (n, %)	132,624 (32.9)
High (n, %)	124,275 (30.8)
Region of complaint	
Spine (n, %)	74,553 (18.5)
Upper extremity (n, %)	109,202 (27.0)
Lower extremity (n, %)	113,449 (28.1)
Other (n, %)	106515 (26.4)
Type of complaint**	
Ankle (n, %)	7,284 (1.8)
Arthritis (n, %)	721 (0.2)
Epicondylitis (n, %)	6,827 (1.7)
Fracture (n, %)	1,245 (0.3)
Hand / wrist (n, %)	23,107 (5.7)
Knee (n, %)	12,308 (3.0)
Knee acute (n, %)	6,460 (1.6)
Low Back Pain (n, %)	20,976 (5.2)
Low Back Pain radicular (n, %)	9,989 (2.5)
Rheumatoid Arthritis (n, %)	520 (0.1)
Shoulder (n, %)	34,579 (8.6)
Others	124,016 (30.7)

Q: Quartile, HCU: High-Cost Users

Chronic disease score: range 0-41. a score of 7 or greater is associated with a 5-fold increase in risk of hospitalization and a 10-fold increase in risk of dying

Q1/Q2* (n=207,625)	Q3 (n= 95,298)	Q4 (n=100,796)	Top 5% HCU (n=20,187)
114,496 (55.1)	53,522 (56.2)	59,083 (58.6)	12,147 (60.2)
51.81 (18.41)	52.63 (18.52)	53.95 (17.82)	54.93 (17.55)
2.56 (3.60)	2.92 (3.76)	3.09 (3.87)	3.20 (3.87)
0.08 (0.31)	0.08 (0.31)	0.08 (0.33)	0.09 (0.33)
1,127 (0.5)	551 (0.6)	547 (0.5)	102 (0.5)
593 (0.3)	364 (0.4)	357 (0.4)	86 (0.4)
807 (0.4)	468 (0.5)	451 (0.4)	91 (0.5)
54,923 (26.5)	27,706 (29.1)	30,350 (30.1)	6,384 (31.6)
10,718 (5.2)	5,696 (6.0)	6344 (6.3)	1,323 (6.6)
78,052 (37.6)	30,528 (32.0)	38,240 (37.9)	7,832 (38.8)
70,775 (34.1)	31,728 (33.3)	30,121 (29.9)	5,767 (28.6)
58,798 (28.3)	33,042 (34.7)	32,435 (32.2)	6,588 (32.6)
36,879 (17.8)	18,156 (19.1)	19,518 (19.4)	5,028 (24.9)
57,007 (27.5)	26,039 (27.3)	26,156 (25.9)	4,737 (23.5)
55,701(26.8)	24,875 (26.1)	32,873 (32.6)	6,543 (32.4)
58,038 (28.0)	26,228 (27.5)	22,249 (22.1)	3,879 (19.2)
4,267 (2.1)	1,628 (1.7)	1,389 (1.4)	213 (1.1)
223 (0.1)	223 (0.2)	275 (0.3)	69 (0.3)
4,383 (2.1)	1,498 (1.6)	946 (0.9)	184 (0.9)
687 (0.3)	358 (0.4)	200 (0.2)	30 (0.1)
11,884 (5.7)	5,518 (5.8)	5,705 (5.7)	866 (4.3)
5,392 (2.6)	3,462 (3.6)	3,454 (3.4)	870 (4.3)
2,761 (1.3)	1,511 (1.6)	2,188 (2.2)	404 (2.0)
10,750 (5.2)	5,318 (5.6)	4,908 (4.9)	1,238 (6.1)
3,554 (1.7)	2,782 (2.9)	3,653 (3.6)	1,040 (5.2)
101 (0.0)	90 (0.1)	329 (0.3)	146 (0.7)
14,895 (7.2)	9,732 (10.2)	9,952 (9.9)	2,374 (11.8)
58,897 (28.3)	32,120 (33.7)	32,999 (32.8)	7,434 (36.8)

**Q 1en Q2 are merged as many patients had the same cost and were therefore hard to disguise.*

** Type of complaint is based on general practitioners' guidelines.[7]

Table 3. Healthcare utilization during 1-year follow-up

	Overall MSK (n=403,719)	**Region Spine (n=74,553)**	**Region Upper extremity (n=109,202)**
Healthcare utilization			
One off consultation (Yes; n, %)	372,743 (92.3)	67,938 (91.1)	102,499 (93.9)
Medication prescribed (Yes; n, %)	68,867 (17.1)	18,169 (24.4)	22,554 (20.7)
Referral (Yes; n, %)	85,042(21.1)	16,132 (21.6)	22,880 (21.0)
Referrals			
Total number of referrals (n)	117,980	22,436	30,074
Referrals to primary care[1] (n, %)	48,424 (41.1)	12,533 (55.9)	12,195 (40.6)
Referrals to secondary care (n, %)	50,636 (42.9)	7,594 (33.8)	14,115 (46.9)
Referrals for imaging (n, %)	8,501 (7.2)	742 (3.3)	1,952 (6.5)
Other referrals[2] (n, %)	10,419 (8.8)	1,567 (7.0)	1,812 (6.0)

MSK; Musculoskeletal compliant, LBP; low back pain

[1] Includes referral to occupational therapy, physiotherapy, exercise therapy, manual therapy, Podotherapy and orthopedic devices, and dietic.

Table 4. Healthcare costs per patient for overall musculoskeletal complaints during 1-year follow-up

	Overall (n=403,719)
Consultations costs (mean, SEM)**	43 (0.03)
Medication costs (mean, SEM)	1 (0.02)
Referrals primary care costs (mean, SEM)	33 (0.14)
Referrals secondary care costs (mean, SEM)	17 (0.07)
Referrals Imaging costs (mean, SEM)	4 (0.03)
Total referrals costs (mean, SEM)	53 (0.18)
Total cost (mean, SEM)	97 (0.18)

Q: Quartile, SEM: Standard Error of the Mean, HCU High-Cost Users

Costs are presented in euros 2021.

**Q 1en Q2 are merged as many patients had the same cost and were therefore hard to distinguish*

Region Lower extremity (n=113,449)	Hand /wrist (n=23,107)	Knee (n=12,308)	LBP (n=20,976)	LBP radicular pain (n=9,989)	Shoulder (n=34,579)
104,144 (91.8)	21,964 (95.1)	11,231 (91.2)	19,229 (91.7)	8,856 (88.7)	31,896 (92.2)
13,568 (12.0)	4,542 (19.7)	3,833 (31.1)	5,908 (28.2)	3,502 (35.1)	11,943 (34.5)
28,103 (24.8)	5,156 (22.3)	2,900 (23.6)	4,108 (19.6)	3,056 (30.6)	8,523 (24.6)
38,495	6,579	4,206	5,386	4,492	11,422
14,451 (37.5)	1,942 (29.5)	1,511 (35.9)	3,468 (64.4)	2,058 (45.8)	6,105 (53.5)
18,214 (47.3)	4,089 (62.2)	2,299 (54.7)	1,468 (27.3)	2045 (45.5)	4,415 (38.7)
3,088 (8.0)	264 (4.0)	201 (4.8)	165 (3.0)	116 (2.6)	395 (3.4)
2,742 (7.2)	284 (4.3)	195 (4.6)	285 (5.3)	273 (6.1)	507 (4.4)

[2] Includes referrals to non pre specified disciplines that deemed to be not clinically relevant as it is possible that patients consult their GP for more than one complaint during the same consultation, while the consultation was assigned to an ICPC-code related to the musculoskeletal complaint.

Q1/Q2* (n=207,625)	Q3 (n= 95,298)	Q4 (n=100,796)	Top 5% HCU (n=20,187)
36 (0.00)	50 (0.06)	51 (0.12)	55 (0.34)
0.00 (0.00)	3 (0.01)	3 (0.08)	6 (1.12)
0.00 (0.00)	2 (0.03)	130 (0.42)	279 (1.10)
0.00 (0.00)	4 (0.04)	63 (0.22)	106 (0.74)
0.00 (0.00)	1 (0.02)	14 (0.11)	20 (0.32)
0.00 (0.00)	7 (0.05)	206 (0.44)	405 (1.11)
36 (0.08)	58 (0.05)	259 (0.44)	466 (1.12)

**Consultation cost refer to GP consultations (*office-based consultations, home visits, consultations by phone/email, and consultations by a GP-based nurse specialist or physician assistant)

Table 5. Overall total healthcare costs per region and type of complaint

	Total healthcare costs (mean, SEM)	Total costs consultation (mean, SEM)	Total cost referral Primary care givers (mean, SEM)
Overall	39,180,531 (0.18)	17,313,539 (0.03)	13,269,305 (0.03)
Region			
Spine	8,027,792 (0.50)	3,283,034 (0.08)	3,466,607 (0.39)
Upper extremity	10,156,463 (0.33)	4,508,261 (0.05)	3,315,373 (0.25)
Lower extremity	11,871,861 (0.36)	4,913,076 (0.07)	3,941,983 (0.26)
Type of complaint			
Hand/wrist	2,079,571 (0.66)	936,386 (0.11)	527,759 (0.46)
Knee	1,315,833 (1.16)	548,362 (0.23)	406,809 (0.80)
Low Back Pain	2,111,707 (0.88)	896,601 (0.13)	955,999 (0.74)
Low Back Pain radicular	1,371,777 (1.58)	464,041 (0.26)	576,642 (1.18)
Shoulder	3,859,345 (0.71)	1,476,802 (0.11)	1,661,514 (0.57)

SEM: Standard error of the mean
Costs are presented in euros 2021.

Discussion

Main findings

Most patients visiting a GP with a musculoskeletal complaint only received a single consultation. The one-year mean annual direct healthcare costs per patient were relatively similar across conditions, except for low-back pain with radiation. The total annual direct healthcare costs in the complete study sample amounted to €39,180,531. The key cost driver consisted of referrals, with total costs of €21,412,524, more than half of which was for referrals to primary care providers. The top 5% of high-cost users were responsible for 24% of the costs. High age, being female, low social economic status, spine complaints, a high number of musculoskeletal diagnoses, and a high comorbidity score were predictive factors for having higher direct healthcare costs, but only explained 0.7% of the variance.

Comparison with literature

The mean annual direct healthcare costs per patient were low compared to that of other conditions treated in general practice. For example, the study by Redekop et al,[28] which included questionnaire-based data from 29 GP-practices on the medical consumption of Dutch patients with diabetes mellitus, found that the average annual healthcare costs for GP-guided care for patients with diabetes mellitus was €600 per patient per year, making care for patients with diabetes mellitus approximately 6 times more expensive.

Total cost referral medical specialist (mean, SEM)	Total cost imaging (mean, SEM)	Total cost referral (mean, SEM)	Total cost medication (mean, SEM)
6,703,684 (0.14)	1,439,536 (0.07)	21,412,524 (0.18)	466.684 (0.02)
998,417 (0.15)	126,503 (0.05)	4,591,527 (0.47)	154,139 (0.08)
1,864,738 (0.13)	331,301 (0.05)	5,511,412 (0.32)	134,221 (0.04)
2,407,213 (0.15)	525,876 (0.06)	6,875,072 (0.35)	91,193 (0.03)
539,132 (0.34)	44,926 (0.08)	1,111,816 (0.65)	29,432 (0.05)
299,247 (0.51)	35,113 (0.14)	741,169 (1.11)	26,390 (0.07)
194,005 (0.22)	28,651 (0.09)	1,178,656 (0.85)	37,468 (0.11)
264,590 (0.60)	21,129 (0.14)	862,362 (1.49)	40,842 (0.34)
581,112 (0.24)	65,816 (0.08)	2,308,442 (0.69)	74,332 (0.11)

Our findings seem to suggest some level of agreement between clinical recommendations and clinical practice in the Netherlands.[8] To illustrate, in accordance with the Dutch guidelines for low back pain and hand/wrist pain, relatively few patients with low back pain and relatively many patients with hand/wrist complaints were referred to secondary care facilities and diagnostic imaging.[9,10] This seems to suggest that Dutch GPs largely adhere to existing GP guidelines, but further research is needed to confirm this.

Although the accurate calculation of population-based estimates requires sample weights, a gross estimate of the total costs of GP guided care for patients with musculoskeletal complaints can be put at €301,3 million for the total Dutch population (i.e., €39,180,531/13*100). This is slightly less than a cost estimate reported by the Dutch Ministry of Health, Welfare and Sports (i.e., €342,5 million).[29] The difference in costs is probably because the Ministry's estimates included a slightly broader range of healthcare resources, but probably most importantly because of its use of tariffs to value healthcare utilization. Tariffs are typically higher than the standard prices used in this study and are generally discouraged by health economic guidelines, such as the Dutch guideline for costing studies, because they can differ extensively across healthcare providers and insurers and do not represent opportunity costs due to the regulated nature of the healthcare market.[13,22]

Our findings also showed that in GP guided care, average primary care referral costs were higher than average secondary care referral costs (i.e., €303 versus €155). This contrasts with how primary and secondary care compare in the total cost of musculoskeletal care, where hospital and medical-specialist care make up almost half of the total cost (i.e., €2,710 of €6,569 million).[29] This difference is partly due to the way healthcare costs were allocated. For primary care referrals, costs were allocated based on treatment episodes (since a GP typically requests treatment in these cases). For secondary care referrals, costs were allocated based on a single consultation (and did not include costs related to diagnostic imaging, hospitalization and surgery since a GP typically requests a consultation rather than follow-up examinations, surgery, or possible admissions).

Similar studies predicting having higher direct healthcare costs of GP guided care are lacking. Nonetheless, a retrospective cohort study[18] using Medical Expenditure Panel Survey data showed that various non-modifiable factors were associated with having higher healthcare costs in patients with musculoskeletal pain. These factors included age, diagnosis type, and number of musculoskeletal diagnoses, which is in line with our findings. However, the retrospective cohort study also found several modifiable factors to be associated with having higher healthcare costs, such as a higher number of missed workdays, and greater pain interference, while higher self-reported physical and mental health were associated with lower healthcare costs. Unfortunately, we were not able to include these modifiable factors in our analysis, as we used clinical registration data which did not include this information. Another study by Killingmo et al,[14] on modifiable prognostic factors of high healthcare costs among older people seeking primary care due to back pain, using data from two cohort studies, reported that a higher level of pain severity, disability, depression, and a lower level of physical health-related quality of life were associated with having higher healthcare costs. Again, since the data used in our study were collected for clinical purposes only, information on these types of modifiable prognostic factors were scarce and therefore only partially included in our analysis.

Strengths and limitations

This study was the first to provide an overview of healthcare utilization and direct healthcare costs of GP-guided care, stratified by specific regions and musculoskeletal conditions. There are several limitations of our study, because of the use of clinical registration data. First, the registration of ICPC codes and healthcare utilization is generally of suboptimal quality due to issues with accuracy, concordance, and plausibility, leading to validity issues.[32] To partially account for this, average cost estimates were not reported by ICPC code but at a more aggregate level (i.e., region

of complaint and type of complaint), and furthermore, in the case of secondary care referrals, we were also able to partially correct for this by not including referrals that are clearly unrelated to a musculoskeletal complaint (e.g. ophthalmologist referral). However, for referrals to primary care and imaging, this distinction could not be made based on the data available in the dataset. Second, incomplete registration may have led to missing values in some of the variables.[32] Although the amount of missing data was relatively low in our sample (i.e., 93% of the observations were complete cases), data were imputed using information from observed data using MICE.[19] Third, data availability is depended on patients being registered at a GP that is connected to the Pharmo database. This means that data may be less available for younger patients or patients changing GP. However, many Dutch people have the same GP as an adult as they did in their childhood. This means that a complete medical history of younger patients is also available since the time of inception of the database and data availability is not related to being eligible for inclusion. Moreover, very few patients change GP. A report by the Patients' Federation shows that only 3% of all patients change GP each year. Therefore, the risk for potential bias for those that are older and those in the database for a longer time is probably very small. Forth, data on societal cost categories, such as patient costs and costs due to productivity losses from paid and unpaid work could not be included in the current cost estimates. Fifth, a restricted number of available predictive factors in combination with a relatively low level of variation in both the predictive factors as well as the outcome variable (i.e., direct healthcare costs) led to the low predictive performance of the regression model with statistically significant, but relatively small betas. However, in a large population, the budget impact can be high despite the betas being small. To illustrate, based on the model's estimates, females have - on average - €6 higher healthcare costs than men. As our sample included 227,101 women resulting in a total cost difference of €1,378,503 (i.e., €6 * 227,101), equaling roughly €10,603,869 in the total Dutch population. Sixth, the use of stepwise selection methods based on significance level are sub-optimal for predictive models. However, a recent overview on the selection of variables of Sauerbrei et al[31] showed that there is not yet enough evidence on which to base recommendations for the selection of variables in multivariable analysis. Therefore, we used stepwise selection as this method already starts with a plausible model. Last, this study provides an overview of the healthcare utilization and direct healthcare cost of GP-guided care amongst Dutch MSK patients, based on a highly representative sample of 13% of all adult patients registered with a Dutch GP practice (n=16,479,000). Please note, however, that the current cost estimates cannot be interpreted as a proxy of the total (primary) healthcare costs of musculoskeletal conditions in the Netherlands, as only the cost of care guided or referred to by GPs was included and, for example, hospitalization, surgeries, and physiotherapy visits without a GP referral were not included.

Conclusion

Most patients had a single consultation and the mean annual direct healthcare costs of GP-guided care in patients with a musculoskeletal condition was relatively low. Healthcare costs did not differ considerably across specific conditions. Key cost driver of GP-guided care in patients with musculoskeletal conditions were GP consultations followed by referrals to primary care (including consultations in primary care). The investigated set of predictive factors explained a negligible part of the variance in direct healthcare costs. Thus, it is unclear which factors do explain high direct healthcare costs in patients with musculoskeletal complaints.

References

1. Barber J, Thompson S. Multiple regression of cost data: use of generalised linear models. J Health Serv Res Policy 2004;9:197-204. doi:10.1258/1355819042250249
2. Blum A, Kalai A, Langford J. Beating the hold-out: Bounds for k-fold and progressive cross-validation. Annual Conference Computational Learning Theory 1999. doi:10.1145/307400.307439
3. Burgess R, Hall J, Bishop A, Lewis M, Hill J. Costing Methodology and Key Drivers of Health Care Costs Within Economic Analyses in Musculoskeletal Community and Primary Care Services: A Systematic Review of the Literature. J Prim Care Community Health 2020;11:1-13. doi:10.1177/2150132719899763
4. van Buuren S, Groothuis-Oudshoorn K. MICE: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software 2011;45:1-67. doi:10.18637/jss.v045.i03
5. Chowdhury MZI, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. Fam Med Community Health 2020;8:e000262. doi:10.1136/fmch-2019-000262
6. Clark DO, Von Korff M, Saunders K, Baluch WM, Simon GE. A chronic disease score with empirically derived weights. Med Care. 1995;33:783-95. doi: 10.1097/00005650-199508000-00004
7. Dutch College of General Practitioners. GPs and the Dutch healthcare system. Accessed March 27, 2022.
8. Dutch College of General Practitioners. NHG-richtlijnen. Accessed March 27, 2022.
9. Dutch College of General Practitioners. NHG-Standaard Aspecifieke lagerugpijn. Accessed March 27, 2022.
10. Dutch College of General Practitioners. NHG-Standaard Hand- en polsklachten. Accessed March 27, 2022.
11. Dutch Healthcare Authority. Zorgsectoren. Accessed March 27, 2022.
12. Flinterman L, Groenewegen P, Verheij R. Zorglandschap en zorggebruik in een veranderende eerste lijn. Utrecht: Nivel, 2018.
13. Kanters TA, Bouwmans CAM, van der Linden N, Tan SS, Hakkaart-van Roijen L. Update of the Dutch manual for costing studies in health care. PLoS One 2017;12:e0187477. doi: 10.1371/journal.pone.0187477
14. Killingmo RM, Chiarotto A, van der Windt DA, Storheim K, Bierma-Zeinstra SMA, Småstuen MC, Zolic-Karlsson Z, Vigdal ØN, Koes BW, Grotle M. Modifiable prognostic factors of high costs related to healthcare utilization among older people seeking primary care due to back pain: an identification and replication study. BMC Health Serv Res 2022 Jun 18;22:793. doi: 10.1186/s12913-022-08180-2.
15. Kort D, van Rein N, van der Meer FJM, Vermaas HW, Wiersma N, Cannegieter SC, Lijfering WM. Relationship between neighborhood socioeconomic status and venous thromboembolism: results from a population-based study. J Thromb Haemost. 2017;15:2352-2360. doi: 10.1111/jth.13868
16. Kuhn, M. Building Predictive Models in R Using the caret Package. Journal of Statistical Software 2008;28:1-26. doi.org/10.18637/jss.v028.i05
17. Lamberts H, Wood M. ICPC: International Classification of Primary Care. Oxford: Oxford University Press;1987.
18. Lentz TA, Harman JS, Marlow NM, Beneciuk JM, Fillingim RB, George SZ. Factors associated with persistently high-cost health care utilization for musculoskeletal pain. PLoS One 2019;14:e0225125. doi:10.1371/journal.pone.0225125
19. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. J Clin Epidemiol 2019;110:63-73. doi:10.1016/j.jclinepi.2019.02.016

20. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BMC Med Res Methodol 2009;9:57. doi:10.1186/1471-2288-9-57
21. Ministry of Public Health, Welfare and Sport. Health care in the Netherlands. Accessed March 27, 2022
22. National Health Care Institute. Guideline for conducting economic evaluations in healthcare [in Dutch: Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg]. Accessed March 27, 2022.
23. National Health Care Institute. Medication cost. Accessed March 27, 2022.
24. Netherlands Institute for Health Services Research. Zorgregistraties eerste lijn. Accessed March 27, 2022.
25. van Pijkeren R, Lekkerkerk T. Meldactie 'Huisartsenzorg'. Utrecht: Patiëntenfederatie Nederland, 2016.
26. Putnam KG, Buist DS, Fishman P, Andrade SE, Boles M, Chase GA, Goodman MJ, Gurwitz JH, Platt R, Raebel MA, Arnold Chan K. Chronic disease score as a predictor of hospitalization. Epidemiology. 2002;13:340-346. doi: 10.1097/00001648-200205000-00016.
27. PHARMO institute. Accessed March 27, 2022.
28. Redekop WK, Koopmanschap MA, Rutten GE, Wolffenbuttel BH, Stolk RP, Niessen LW. Resource consumption and costs in Dutch patients with type 2 diabetes mellitus. Results from 29 general practices. Diabet Med. 2002;19:246-253. doi:10.1046/j.1464-5491.2002.00654.x
29. Ministry of Public Health, Welfare and Sport. Kosten van ziekten. Accessed March 27, 2022.
30. Ministry of Public Health, Welfare and Sport. Zorguitgaven. Accessed March 27, 2022.
31. Sauerbrei W, Perperoglou A, Schmid M, Abrahamowicz M, Becher H, Binder H, Dunkler D, Harrell FE Jr, Royston P, Heinze G; for TG2 of the STRATOS initiative. State of the art in selection of variables and functional forms in multivariable analysis-outstanding issues. Diagn Progn Res. 2020;4:3. doi: 10.1186/s41512-020-00074-3.
32. Sollie JW, Numans ME, Helsper CW. Zijn de gegevens uit uw HIS een goudmijn? Huisarts Wet 2018;61. doi:10.1007/s12445-018-0005-7.
33. The Netherlands Institute for Social Research. Neighborhood Social Economic Status. Accessed June 9th, 2018).
34. Trimbos Institute of Mental Health. GGZ in tabellen. Accessed March 27, 2022.
35. Veldkamp R, Meijer, W. Zorg door de fysiotherapeut. Nivel Zorgregistraties Eerste Lijn, Jaarcijfers 2021 en trendcijfers 2018-2021. Utrecht: Nivel, 2022.
36. Wammes JJG, van der Wees PJ, Tanke MAC, Westert GP, Jeurissen PPT. Systematic review of high-cost patients' characteristics and healthcare utilisation. BMJ Open 2018;8:e023113. doi:10.1136/bmjopen-2018-023113
37. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011;30:377-399. doi:10.1002/sim.4067
38. WHO Collaborating Centre for Drug Statistics Methodology. Anatomical Therapeutic Chemical (ATC) classification system. Accessed March 27, 2022.

Appendix I. Included ICPC codes

ICPC code	Description	Total 403,719 (n, %)	Region	Type of compliant
L01	Neck symptoms/complaints [ex. N01]	17,036 (4.2)	1	0
L02	Back symptoms/complaints	21,076 (5.2)	1	0
L03	Low back pain without radiation [ex. L86]	20,976 (5.2)	1	1
L04	Chest symptoms/complaints	26,652 (6.6)	0	0
L05	Flank symptoms/complaints	3,889 (1.0)	0	0
L06	Armpit symptoms/problems	1,398 (0.3)	0	0
L07	Jaw (joint) symptoms/complaints	4,291 (1.1)	0	0
L08	Shoulder symptoms/problems	25,317 (6.3)	2	8
L09	Arm symptoms/problems	6,852 (1.7)	2	0
L10	Elbow symptoms/problems	4,048 (1.0)	2	0
L11	Wrist symptoms/problems	7,260 (1.8)	2	0
L12	Hand/finger symptoms/conditions	18,346 (4.5)	2	0
L13	Hip symptoms/conditions	9,986 (2.5)	3	0
L14	Leg/thigh symptoms/conditions	17,478 (4.3)	3	0
L15	Knee symptoms/conditions	24,130 (6.0)	3	0
L16	Ankle symptoms/conditions	5,635 (1.4)	3	0
L17	Foot/toe symptoms/problems	25,840 (6.4)	3	0
L17.01	Metatarsalgia	967 (0.2)	3	0
L18	Muscle pain	6,633 (1.6)	0	0
L18.01	Fibromyalgia	386 (0.1)	0	0
L19	Symptoms multiple/nonspecific muscles	4,353 (1.1)	0	0
L20	Symptoms multiple/unspecified joints	4,087 (1.0)	0	0
L26	Fear of musculoskeletal cancer	10 (0.0)	0	0
L27	Fear of other musculoskeletal disease	98 (0.0)	0	0
L28	Musculoskeletal function limitation/disability	508 (0.1)	0	0
L29	Other complaints musculoskeletal system	4,137 (1.0)	0	0
L44	Preventive medication	75 (0.0)	0	0
L49	Prevention	30 (0.0)	0	0
L49.01	Fracture prevention	98 (0.0)	0	11
L70	Infectious musculoskeletal disease	90 (0.0)	0	0
L70.01	Osteomyelitis	50 (0.0)	0	0
L70.02	Septic arthritis	30 (0.0)	0	9
L71	Neoplasm musculoskeletal system	140 (0.0)	0	0
L71.01	Musculoskeletal malignancy	33 (0.0)	0	0

ICPC code	Description	Total 403,719 (n, %)	Region	Type of compliant
L71.02	Benign musculoskeletal neoplasm	117 (0.0)	0	0
L72	Fracture of radius/ulna	2,613 (0.6)	2	0
L73	Fracture of tibia/fibula	1,730 (0.4)	3	0
L74	Hand/foot fracture	2,932 (0.7)	0	0
L74.01	Fracture ossa phalanges hand	1,014 (0.3)	2	0
L74.02	Fracture ossa phalanges foot	1,107 (0.3)	3	0
L75	Fracture femur	785 (0.2)	3	0
L75.01	Fracture collum femoris	630 (0.2)	3	0
L76	Other fracture	930 (0.2)	0	0
L76.01	Fracture skull	125 (0.0)	0	0
L76.02	Fracture of nose	280 (0.1)	0	0
L76.03	Clavicle fracture	650 (0.2)	2	0
L76.04	Humerus fracture	860 (0.2)	2	0
L76.05	Rib fracture	786 (0.2)	0	0
L76.06	Fracture of vertebral column	716 (0.2)	1	0
L76.07	Pelvic fracture	253 (0.1)	3	0
L76.08	Fracture patella	93 (0.0)	3	0
L77	Sprain/distortion of ankle	7,284 (1.8)	3	5
L78	Knee sprain/distortion	3,836 (1.0)	3	4
L79	Other sprain/distortion	999 (0.2)	0	0
L79.01	Whiplash trauma to cervical spine	175 (0.0)	1	0
L80	Luxation/subluxation	473 (0.1)	0	0
L80.01	(Sub)luxation shoulder	401 (0.1)	2	0
L80.02	(Sub)luxation jaw joint	55 (0.0)	0	0
L80.03	(Sub)luxation finger	173 (0.0)	2	0
L80.04	(Sub)luxation acromio-clavicular joint	163 (0.0)	2	0
L80.05	(Sub)luxation radius head/sun arm	7 (0.0)	2	0
L81	Other musculoskeletal injuries	12,785 (3.2)	0	0
L81.01	Coup de fouet / whiplash	1,642 (0.4)	0	0
L81.02	Rib contusion	4,016 (1.0)	0	0
L82	Congenital musculoskeletal abnormality(s)	147 (0.0)	0	0
L82.01	Congenital hip luxation/hip dysplasia	29 (0.0)	0	0
L82.02	Spina bifida occulta	2 (0.0)	0	0
L82.03	Neck rib	4 (0.0)	0	0
L82.04	Clubfoot	13 (0.0)	0	0
L83	Syndrome cervical spine	1,645 (0.4)	1	0

ICPC code	Description	Total 403,719 (n, %)	Region	Type of compliant
L83.01	Hernia cervicalis	364 (0.1)	1	0
L84	Osteoarthritis/spondylosis vertebral column	927 (0.2)	1	0
L84.01	Osteoarthritis/spondylosis	571 (0.1)	1	0
L84.02	Spondylolysis/listhesis	53 (0.0)	1	0
L85	Acquired disorder(s) of the spine	197 (0.0)	1	0
L85.01	Scoliosis	259 (0.1)	1	0
L86	Low back pain with radiation	9,989 (2.5)	1	2
L86.01	HNP (thoracic/lumbar)	1,452 (0.4)	1	0
L87	Ganglion joint/tendon	3,962 (1.0)	2	7
L88	Rheumatoid arthritis/related condition(s)	691 (0.2)	0	9
L88.01	Rheumatoid arthritis	236 (0.1)	0	0
L88.02	Morbus Bechterew (ankylopoetic spondylitis)	75 (0.0)	0	0
L89	Cox osteoarthritis	2,899 (0.7)	3	0
L90	Gonarthrosis	4,769 (1.2)	3	3
L91	Other osteoarthritis/related disease	4,045 (1.0)	2	7
L92	Shoulder syndrome/PHS	9,262 (2.3)	2	8
L93	Epicondylitis lateralis	6,098 (1.5)	2	6
L94	Osgood-Schlatter/other osteochondropathy	51 (0.0)	0	0
L94.01	Osteochondritis dissecans	6 (0.0)	0	0
L94.02	Osgood-Schlatter's disease	18 (0.0)	0	0
L94.03	Epiphysiolysis femoral head	2 (0.0)	0	0
L94.04	Legg-Calvé-Perthes disease	3 (0.0)	0	0
L95	Osteoporosis	1,245 (0.3)	0	11
L95.01	Osteopenia	1,183 (0.3)	0	0
L95.02	Osteoporosis	352 (0.1)	0	0
L96	Acute meniscus/knee ligament injury	845 (0.2)	3	4
L96.01	Meniscus tear (lat./med.)	626 (0.2)	3	4
L96.02	Lig. cruciata injury (anterior/posterior)	187 (0.0)	3	4
L96.03	Collateral knee ligament injury	173 (0.0)	3	4
L96.04	Combined injury menisci/lig.cruciata/ collateral/capillary ligament	33 (0.0)	3	4
L97	Chronic internal trauma to knee	287 (0.1)	3	3
L97.01	Old meniscus injury	41 (0.0)	3	3
L97.02	Non-traumatic meniscal defect	42 (0.0)	3	3
L97.03	Unstable knee	27 (0.0)	3	3
L97.04	Corpus liberum knee	11 (0.0)	3	3

ICPC code	Description	Total 403,719 (n, %)	Region	Type of compliant
L98	Acquired abnormality(s) extremities	562 (0.1)	0	0
L98.01	Mallet finger	651 (0.2)	2	7
L98.02	Pes planus	817 (0.2)	3	0
L98.03	Hallux valgus	1,517 (0.4)	3	0
L98.04	Hammer toe	315 (0.1)	3	0
L98.05	Leg length difference	92 (0.0)	3	0
L99	Other musculoskeletal disease(s)	7,993 (2.0)	0	0
L99.01	Bursitis [ex. L92]	6,546 (1.6)	0	0
L99.02	Tendovaginitis/tendinitis	6,297 (1.6)	2	7
L99.03	Dupuytren's contracture	1,273 (0.3)	2	7
L99.04	Trigger finger	2,800 (0.7)	2	7
L99.05	Epicondylitis medialis	729 (0.2)	2	6
L99.06	Tietze's syndrome	1,030 (0.3)	0	0
L99.07	Retropatellar chondropathy/patellofemoral syndrome	975 (0.2)	3	3
L99.08	Heel spur/plantar fasciitis	5,017 (1.2)	3	0
L99.09	Hyperlaxity	54 (0.0)	0	0
L99.10	Corpus liberum joint [ex. L97]	6 (0.0)	0	0
L99.11	Pseudarthrosis	7 (0.0)	0	0
L99.12	Polymyalgia rheumatica	520 (0.1)	0	10
L99.13	Artritis psoriatica	53 (0.0)	0	0
N93	Carpal tunnel syndrome	4,079 (1.0)	2	7

Region: 0= others; 1= spine; 2= upper extremity; 3=lower extremity

Type of complaint: 0= none; 1= Non-specific low back pain; 2= Lumbosacral radicular syndrome; 3= Non-traumatic knee pain; 4= Traumatic knee pain; 5= Ankle ligament injury; 6= Epicondylitis; 7= Hand and wrist pain; 8= Shoulder pain; 9= Arthritis; 10= Polymyalgia rheumatica and arteritis temporalis; 11= Fracture prevention.

Appendix II. Predictive variables

Comorbidities	
Condition	*ICPC codes included*
Depression	P03 Feeling depressed P76 Depressive disorder
Cardiovasculair disease	K74 Ischaemic heart disease w. angina K75 Acute myocardial infarction K76 Ischaemic heart disease w/o angina K77 Heart failure K85 Elevated blood pressure K86 Hypertension uncomplicated K87 Hypertension complicated K89 Transient cerebral ischaemia K90 Stroke/cerebrovascular accident K92 Atherosclerosis/PVD K99 Cardiovascular disease other T 93 Lipid disorder
DM	T90.1 Diabetes insulin dependent T90.2 Diabetes non-insulin dependent
Smoking	P17 Tobacco abuse
Obese	T82 Obesity (QI ≥30) T83 Overweight (27≤ QI <30)
Social Economic Status	
SES	The variable Social Economic Status (SES) was determined as a relative measure based on the scores of the Netherlands Institute for Social Research (SCP score). This SCP score is available for all 4-digit postal codes with more than 100 households and is provided every 4 years by the Netherlands Institute for Social Research since 1995. The score is based on mean household income, percentage of households with a low income, percentage of inhabitants without a paid job, and percentage of households with a low mean education. This information is obtained via phone calls from the organization Evers Direct Marketing Besloten Vennootschap (EDM-BV) to 1 person in each 6-digit postal code (usually 1 street) and aggregated to 4-digit postal codes. To determine 4 categories of SES in our cohort (low, middle, high, unknown), we assigned the most recent SCP scores to all patients in the PHARMO Database Network and used tertiles as cutpoints for classification to low, middle and high. Patients without an SCP code available were classified as unknown.

Appendix III. Baseline characteristics per region

Spine	**Overall (n= 74,553)**
Gender; female (Yes; n, %)	42,451 (56.9)
Age (mean, SD)	51.47 (18.23
Comorbidities	
Chronic Disease Score (mean, SD)	2.63 (3.62)
Number of MSK diagnosis (mean, SD)	0.07 (0.30)
Depression (Yes; n, %)	431 (0.6)
Obese (Yes; n, %)	231 (0.3)
Smoking (Yes; n, %)	321 (0.4)
Cardiovascular (Yes; n, %)	19,776 (26.5)
Diabetes (Yes; n, %)	3,989 (5.4)
Social Economic Status	
Low (n, %)	27,099 (36.3)
Middle (n, %)	24,238 (32.5)
High (n, %)	23,216 (31.1)
Upper extremity	**Overall (n=109,202)**
Gender; female (Yes; n, %)	60,762 (55.6)
Age (mean, SD))	52.68 (17.49)
Comorbidities	
Chronic Disease Score (mean (SD))	2.68 (3.63)
Number of MSK diagnosis (mean (SD))	0.08 (0.31)
Depression (Yes; n, %)	571 (0.5)
Obese (Yes; n, %)	335 (0.3)
Smoking (Yes; n, %)	479 (0.4)
Cardiovascular (Yes; n, %)	30,179 (27.6)
Diabetes (Yes; n, %)	6,178 (5.7)
Social Economic Status	
Low (n, %)	39,439 (36.1
Middle (n, %)	36,086 (33.0)
High (n, %)	33,677 (30.8)

Q1/Q2* (n=37,277)	Q3 (n=18,638)	Q4 (n=18,638)	Top 5% HCU (n=3,728)
20,791 (55.8)	10,663 (57.2)	10,997 (59.0)	2,187 (58.7)
50.89 (18.20)	51.44 (18.18)	52.64 (18.29)	53.19 (18.51)
2.38 (3.46)	2.97 (3.70)	2.97 (3.81)	3.13 (3.89)
0.07 (0.29)	0.08 (0.31)	0.08 (0.30)	0.08 (0.31)
202 (0.5)	118 (0.6)	111 (0.6)	20 (0.5
93 (0.2)	81 (0.4)	57 (0.3)	18 (0.5)
137 (0.4)	108 (0.6)	76 (0.4)	15 (0.4)
9,334 (25.0)	5,115 (27.4	5,327 (28.6)	1,107 (29.7)
1,767 (4.7)	1,082 (5.8)	1,140 (6.1)	240 (6.4)
14,258 (38.2)	5,810 (31.2)	7,031 (37.7)	1,379 (37.0)
12,535 (33.6)	6,119 (32.8)	5,584 (30.0)	1,026 (27.5)
10,484 (28.1)	6,709 (36.0)	6,023 (32.3)	1,323 (35.5)
Q1/Q2 (n=57,007)	**Q3 (n=25,075)**	**Q4 (n=27,120)**	**Top 5% HCU (n=5,552)**
31,223 (54.8)	13,931 (55.6)	15,608 (57.6)	3,282 (59.1)
52.01 (17.66)	53.04 (17.51)	53.78 (17.03)	54.17 (16.41)
2.46 (3.51)	2.84 (3.69)	2.98 (3.79)	3.07 (3.81)
0.08 (0.31)	0.08 (0.31)	0.09 (0.33)	0.08 (0.32)
291 (0.5)	133 (0.5)	147 (0.5)	31 (0.6)
156 (0.3)	90 (0.4)	89 (0.3)	20 (0.4)
220 (0.4)	130 (0.5)	129 (0.5)	28 (0.5)
14,919 (26.2)	7,339 (29.3)	7,921 (29.2)	1,733 (31.2)
2,908 (5.1)	1,549 (6.2)	1,721 (6.3)	383 (6.9)
21,306 (37.4)	7,953 (31.7)	10,180 (37.5)	2,278 (40.1)
19,511 (34.2)	8,374 (33.4)	8,201 (30.2)	1,597 (28.1)
16,190 (28.4)	8,748 (34.9)	8,739 (32.2)	1,800 (31.7)

Lower extremity	**Overall (n=113,449)**
Gender; female (Yes; n, %)	64,652 (57.0)
Age (mean, SD)	53.61 (19.07)
Comorbidities	
Chronic Disease Score (mean, SD)	2.91 (3.80)
Number of MSK diagnosis (mean, SD)	0.08 (0.31)
Depression (Yes; n, %)	584 (0.5)
Obese (Yes; n, %)	376 (0.3)
Smoking (Yes; n, %)	413 (0.4)
Cardiovascular (Yes; n, %)	33,152 (29.2)
Diabetes (Yes; n, %)	6,711 (5.9)
Social Economic Status	
Low (n, %)	41,492 (36.6)
Middle (n, %)	37,001 (32.6)
High (n, %)	34,956 (30.8)

**Q 1en Q2 are merged as many patients had the same cost and were therefore hard to disguise.*

Appendix IV. Baseline characteristics per type of complaint

Hand/wrist	**Overall (n=23,107)**
Gender; female (Yes; n, %)	13,925 (60.3)
Age (mean, SD)	56.69 (16.44)
Comorbidities	
Chronic Disease Score (mean, SD)	3.09 (3.79)
Number of MSK diagnosis (mean, SD)	0.09 (0.32)
Depression (Yes; n, %)	114 (0.5)
Obese (Yes; n, %)	69 (0.3)
Smoking (Yes; n, %)	99 (0.4)
Cardiovascular (Yes; n, %)	7,631 (33.0)
Diabetes (Yes; n, %)	1,593 (6.9)
Social Economic Status	
Low (n, %)	8,361 (36.2)
Middle (n, %)	7,801 (33.8)
High (n, %)	6,945 (30.1)

	Q1/Q2* (n=56,756)	Q3 (n=28,338)	Q4 (n=28,355)	Top 5% HCU (n=5,675)
	31,919 (56.2)	16,095 (56.8)	16,638 (58.7)	3,449 (60.8)
	52.50 (19.26)	54.23 (19.41)	55.20 (18.19)	57.29 (17.48)
	2.65 (3.67)	3.21 (3.94)	3.14 (3.89)	3.33 (3.90)
	0.07 (0.30)	0.08 (0.31)	0.08 (0.32)	0.09 (0.34)
	298 (0.5)	152 (0.5)	134 (0.5)	27 (0.5)
	163 (0.3)	102 (0.4)	111 (0.4)	22(0.4)
	186 (0.3)	119 (0.4)	108 (0.4)	19 (0.3)
	15,439 (27.2)	8,816 (31.1)	8,897 (31.4)	1,873 (33.0)
	3,163 (5.6)	1,798 (6.3)	1,750 (6.2)	372 (6.6)
	20,986 (37.0)	9,375 (33.1)	11,131(39.3)	2,278 (40.1)
	19,314 (34.0)	9,419 (33.2)	8,268 (29.2)	1,597 (28.1)
	16,456 (29.0)	9,544 (33.7)	8,956 (31.6)	1,800 (31.7)

	Q1/Q2* (n=11,884)	Q3 (n=5,447)	Q4 (n=5,776)	Top 5% HCU (n=1,335)
	7,127 (60.0)	3,242 (59.5)	3,556 (61.6)	836 (62.6)
	55.94 (16.82)	57.85 (16.39)	57.14 (15.58)	55.72 (15.46)
	2.85 (3.67)	3.36 (3.91)	3.33 (3.87)	3.15 (3.77)
	0.08 (0.32)	0.08 (0.31)	0.09 (0.34)	0.09 (0.32)
	57 (0.5	23 (0.4)	34 (0.6)	6 (0.4)
	32 (0.3)	18 (0.3)	19 (0.3)	1 (0.1)
	43 (0.4)	23 (0.4)	33 (0.6)	9 (0.7)
	3,756 (31.6)	1,946 (35.7)	1,929 (33.4)	449 (33.6)
	755 (6.4)	393 (7.2)	445 (7.7)	110 (8.2)
	4,482 (37.7)	1,778 (32.6)	2,101 (36.4)	509 (38.1)
	4,119 (34.7)	1,892 (34.7)	1,790 (31.0)	425 (31.8)
	3,283 (27.6)	1,777 (32.6)	1,885 (32.6)	401 (30.0)

Knee	**Overall (n=12,308)**
Gender; female (Yes; n, %)	6,471 (52.6)
Age (mean, SD)	60.57 (16.13)
Comorbidities	
Chronic Disease Score (mean, SD)	3.66 (4.02)
Number of MSK diagnosis (mean, SD)	0.09 (0.33)
Depression (Yes; n, %)	56 (0.5)
Obese (Yes; n, %)	43 (0.3)
Smoking (Yes; n, %)	45 (0.4)
Cardiovascular (Yes; n, %)	4,596 (37.3)
Diabetes (Yes; n, %)	967 (7.9)
Social Economic Status	
Low (n, %)	4,273 (34.7)
Middle (n, %)	4,372 (35.5)
High (n, %)	3,663 (29.8)
Low Back Pain	**Overall (n=20,976)**
Gender; female (n, %)	11,292 (53.8)
Age (mean, SD)	50.96 (18.13)
Comorbidities	
Chronic Disease Score (mean, SD)	2.46 (3.51)
Number of MSK diagnosis (mean, SD)	0.06 (0.28)
Depression (Yes; n, %)	129 (0.6)
Obese (Yes; n, %)	60 (0.3)
Smoking (Yes; n, %)	91 (0.4)
Cardiovascular (Yes; n, %)	5,384 (25.7)
Diabetes (Yes; n, %)	1,090 (5.2)
Social Economic Status	
Low (n, %)	7,570 (36.1)
Middle (n, %)	6,968 (33.2)
High (n, %)	6,438 (30.7)

Q1/Q2* (n=6,187)	Q3 (n=3,047)	Q4 (n=3,074)	Top 5% HCU (n=616)
2,846 (46.0)	1,678 (55.1	1,947 (63.3)	423 (68.7)
57.95 (16.80)	63.27 (15.73)	63.17 (14.16	63.41 (13.44)
3.19 (3.88)	4.15 (4.12)	2.03 (2.97)	4.00 (4.07)
0.08 (0.31)	0.09 (0.35)	0.10 (0.36)	0.12 (0.42)
27 (0.4)	11 (0.4)	18 (0.6)	4 (0.6)
19 (0.3)	11 (0.4)	13 (0.4)	4 (0.6)
28 (0.5)	13 (0.4)	4 (0.1)	0 (0.0)
2,073 (33.5)	1,271 (41.7)	1,252 (40.7)	253 (41.1)
416 (6.7)	253 (8.3)	298 (9.7)	63 (10.2)
2,262 (36.6)	912 (29.9)	1,099 (35.8)	253 (41.1)
2,227 (36.0)	1,161 (38.1)	984 (32.0)	169 (27.4)
1,698 (27.4)	974 (32.0)	991 (32.2)	194 (31.5)
Q1/Q2* (n=10,750)	**Q3 (n=4,982)**	**Q4 (n=5,244)**	**Top 5% HCU (n=1,061)**
5,646 (52.5)	2,635 (52.9	3,011 (57.4)	621 (58.5)
50.90 (17.68)	50.26 (18.17)	51.74 (18.95)	52.92 (19.53)
2.26 (3.34)	2.49 (3.46)	2.86 (3.82)	2.98 (3.74)
0.06 (0.27)	0.07 (0.29)	0.07 (0.30)	0.07 (0.28)
59 (0.5)	34 (0.7)	36 (0.7)	4 (0.4)
27 (0.3)	19 (0.4)	14 (0.3)	7 (0.7)
36 (0.3)	33 (0.7)	22 (0.4)	4 (0.4)
2,629 (24.5)	1,285 (25.8)	1,470 (28.0)	318 (30.0)
502 (4.7)	263 (5.3)	325 (6.2)	74 (7.0)
4,087 (38.0)	1,497 (30.0)	1,986 (37.9)	400 (37.7)
3,725 (34.7)	1,661 (33.3)	1,582 (30.2)	306 (28.8)
2,938 (27.3)	1,824 (36.6)	1,676 (32.0)	355 (33.5)

Low Back Pain radicular pain	**Overall (n=9,989)**
Gender; female (Yes; n, %)	5,722 (57.3)
Age (mean, SD)	55.03 (16.72)
Comorbidities	
Chronic Disease Score (mean, SD)	3.02 (3.78)
Number of MSK diagnosis (mean, SD)	0.08 (0.31)
Depression (Yes; n, %)	49 (0.5)
Obese (Yes; n, %)	39 (0.4
Smoking (Yes; n, %)	51 (0.5)
Cardiovascular (Yes; n, %)	3,060 (30.6)
Diabetes (Yes; n, %)	637 (6.4)
Social Economic Status	
Low (n, %)	3,402 (34.1)
Middle (n, %)	3,291 (32.9)
High (n, %)	3,296 (33.0)
Shoulder	**Overall (n=34,579)**
Gender; female (Yes; n, %)	18,495 (53.5)
Age (mean, SD)	53.98 (16.82)
Comorbidities	
Chronic Disease Score (mean, SD)	2.78 (3.67)
Number of MSK diagnosis (mean, SD)	0.08 (0.31)
Depression (Yes; n, %)	189 (0.5)
Obese (Yes; n, %)	124 (0.4)
Smoking (Yes; n, %)	153 (0.4)
Cardiovascular (Yes; n, %)	9,841 (28.5)
Diabetes (Yes; n, %)	2117 (6.1)
Social Economic Status	
Low (n, %)	12,539 (36.3)
Middle (n, %)	11,603 (33.6)
High (n, %)	10,437 (30.2)

**Q 1en Q2 are merged as many patients had the same costs and were therefore hard to distinguish*

Q1/Q2* (n=4,995)	Q3 (n=2,500)	Q4 (n=2,494)	Top 5% HCU (n=500)
2,789 (55.8)	1,472 (58.9)	1,461 (58.6)	312 (62.4)
54.34 (16.46)	56.15 (16.90)	55.29 (16.99)	54.27 (17.34)
2.70 (3.61)	3.60 (4.04)	3.10 (3.78)	3.25 (3.88)
0.07 (0.29)	0.09 (0.34)	0.08 (0.30	0.09 (0.29)
19 (0.4)	14 (0.6)	16 (0.6)	1 (0.2)
15 (0.3)	14 (0.6)	10 (0.4)	3 (0.6)
30 (0.6)	11 (0.4)	10 (0.4)	5 (1.0)
1,416 (28.3)	852 (34.1)	792 (31.8)	151 (30.2)
305 (6.1)	170 (6.8)	162 (6.5	32 (6.4)
1,724 (34.5)	802 (32.1)	876 (35.1)	186 (37.2)
1,742 (34.9)	786 (31.4)	763 (30.6)	136 (27.2)
1,529 (30.6)	912 (36.5)	855 (34.3)	178 (35.6)
Q1/Q2* (n=17,349)	**Q3 (n=8,663)**	**Q4 (n= 8,567)**	**Top 5% HCU (n=1,734)**
9,124 (52.6)	4,717 (54.4)	4,654 (54.3)	993 (57.3)
53.46 (16.89)	54.11 (17.35)	54.91 (16.09)	55.56 (15.87)
2.52 (3.50)	3.06 (3.84)	3.00 (3.79)	3.18 (3.79)
0.07 (0.30)	0.08 (0.32)	0.08 (0.33)	0.09 (0.34)
85 (0.5)	51 (0.6)	53 (0.6)	16 (0.9)
57 (0.3)	34 (0.4)	33 (0.4)	7 (0.4)
68 (0.4)	46 (0.5)	39 (0.5)	7 (0.4)
4,721 (27.2)	2,603 (30.0)	2,517 (29.4)	545 (31.4)
986 (5.7)	606 (7.0)	525 (6.1)	124 (7.2)
6,398 (36.9)	2,794 (32.3)	3,347 (39.1)	689 (39.7)
6,072 (35.0)	2,913 (33.6)	2,618 (30.6)	511 (29.5)
4,879 (28.1)	2,956 (34.1)	2,602 (30.4)	534 (30.8)

Appendix V. Costs per region and type of complaint

Spine	Overall (n= 74,553)
Consultation costs (mean, SEM)	44 (0.08)
Medication costs (mean, SEM)	2 (0.08)
Referrals primary care costs (mean, SEM)	47 (0.39)
Referrals secondary care costs (mean, SEM)	13 (0.15)
Imaging costs (mean, SEM)	2 (0.05)
Total referral costs (mean, SEM)	62(0.47)
Total cost (mean, SEM)	93(0.49)
Upper extremity	Overall (n=109,202)
Consultation costs (mean, SEM)	41 (0.05)
Medication costs (mean, SEM)	1 (0.04)
Referrals primary care costs (mean, SEM)	30 (0.25)
Referrals secondary care costs (mean, SEM)	17 (0.13)
Imaging costs (mean, SEM)	3 (0.05)
Total referral costs (mean, SEM)	50 (0.32)
Total cost (mean, SEM)	93 (0.33)
Lower extremity	Overall (n=113,449)
Consultation costs (mean, SEM)	43 (0.07)
Medication costs (mean, SEM)	1 (0.03)
Referrals primary care costs (mean, SEM)	35 (0.26)
Referrals secondary care costs (mean, SEM)	21 (0.15)
Imaging costs (mean, SEM)	5 (0.06)
Total referral costs (mean, SEM)	61 (0.35)
Total cost (mean, SEM)	105 (0.36)
Hand/wrist	Overall (n=23,107)
Consultation costs (mean, SEM)	41 (0.11)
Medication costs (mean, SEM)	1 (0.05)
Referrals primary care costs (mean, SEM)	23 (0.46)
Referrals secondary care costs (mean, SEM)	23 (0.34)
Imaging costs (mean, SEM)	2 (0.08)
Total referral costs (mean, SEM)	48 (0.65)
Total cost (mean, SEM)	90 (0.66)

Q1/Q2* (n=37,277)	Q3 (n=18,638)	Q4 (n=18,638)	Top 5% HCU (n=3,728)
36 (0.00)	52 (0.15)	52 (0.28)	57 (0.83)
0 (0.00)	2 (0.04)	5 (0.32)	16 (1.49)
0 (0.00)	2 (0.08)	184 (1.07)	349 (1.07)
0 (0.00)	3 (0.07)	51 (0.52)	98 (1.75)
0 (0.00)	0 (0.02)	7 (0.20)	12 (0.61)
0 (0.00)	5 (0.11)	241 (1.12)	459 (2.81)
36 (0.00)	59 (0.14)	298 (1.13)	532 (2.91)
Q1/Q2* (n=57,007)	Q3 (n=25,075)	Q4 (n=27,120)	Top 5% HCU (n=5,552)
36 (0.00)	47 (0.10)	47 (0.18)	49 (0.45)
0 (0.00)	2 (0.03)	2 (0.15)	5 (0.71)
0 (0.00)	1 (0.05)	121 (0.79)	270 (1.91)
0 (0.00)	4 (0.08)	65 (0.42)	94 (1.38)
0 (0.00)	0 (0.02)	12 (0.19)	17 (0.56)
0 (0.00)	6 (0.09)	198 (0.80)	380 (1.89)
36 (0.00)	55 (0.09)	247 (0.78)	434 (1.93)
Q1/Q2* (n=56,756)	Q3 (n=28,338)	Q4 (n=28,355)	Top 5% HCU (n=5,675)
36 (0.00)	53 (0.14)	48 (0.20)	54 (0.62)
0 (0.00)	1 (0.03)	2 (0.10)	4 (0.43)
0 (0.00)	4 (0.08)	135 (0.78)	274 (1.99)
0 (0.00)	10 (0.12)	75 (0.44)	119 (1.40)
0 (0.00)	2 (0.05)	17 (0.22)	24 (0.64)
0 (0.00)	16 (0.16)	227 (0.77)	417 (1.94)
36 (0.00)	70 (0.15)	277 (0.77)	475 (1.88)
Q1/Q2* (n=11,884)	Q3 (n=5,447)	Q4 (n=5,776)	Top 5% HCU (n=1,335)
36 (0.00)	45 (0.21)	46 (0.36)	45 (0.77)
0 (0.01)	3 (0.06)	2 (0.21)	3 (0.79)
(0.00)	1 (0.09)	91 (1.54)	243 (3.39)
0 (0.00)	7 (0.21)	87 (0.92)	91 (2.87)
0 (0.00)	0 (0.03)	8 (0.32)	11 (0.80)
0 (0.00)	8 (0.23)	185 (1.53)	346 (3.12)
36 (0.01)	56 (0.22)	232 (1.49)	394 (3.10)

Knee	Overall (n=12,308)
Consultation costs (mean, SEM)	45 (0.23)
Medication costs (mean, SEM)	2 (0.07)
Referrals primary care costs (mean, SEM)	33 (0.80)
Referrals secondary care costs (mean, SEM)	24 (0.51)
Imaging costs (mean, SEM)	3 (0.14)
Total referral costs (mean, SEM)	60 (1.11)
Total costs (mean, SEM)	107 (1.16)
Low Back Pain	Overall (n=20,976)
Consultation costs (mean, SEM)	43 (0.13)
Medication costs (mean, SEM)	2 (0.11)
Referrals primary care costs (mean, SEM)	46 (0.74)
Referrals secondary care costs (mean, SEM)	9 (0.22)
Imaging costs (mean, SEM)	1 (0.09)
Total referral costs (mean, SEM)	56 (0.85)
Total cost (mean, SEM)	101 (0.88)
Low Back Pain radicular pain	Overall (n=9,989)
Consultation costs (mean, SEM)	46 (0.26)
Medication costs (mean, SEM)	4 (0.34)
Referrals primary care costs (mean, SEM)	58 (1.18)
Referrals secondary care costs (mean, SEM)	26 (0.60)
Imaging costs (mean, SEM)	2 (0.14)
Total referral costs (mean, SEM)	86 (1.49)
Total cost (mean, SEM)	137 (1.58)
Shoulder	Overall (n=34,579)
Consultation costs (mean, SEM)	43 (0.11)
Medication costs (mean, SEM)	2 (0.11)
Referrals primary care costs (mean, SEM)	48 (0.57)
Referrals secondary care costs (mean, SEM)	17 (0.24)
Imaging costs (mean, SEM)	2 (0.08)
Total referral costs (mean, SEM)	67 (0.70)
Total cost (mean, SEM)	112 (0.71)

**Q 1en Q2 are merged as many patients had the same costs and were therefore hard to distinguish.*
Costs are presented in euros 2021.

Q1/Q2* (n=6,187)	Q3 (n=3,047)	Q4 (n=3,074)	Top 5% HCU (n=616)
36 (0.00)	52 (0.40)	54 (0.79)	61 (2.47)
0 (0.02)	3 (0.12)	4 (0.23)	4 (0.48)
0 (0.00)	2 (0.18)	130 (2.45)	293 (6.00)
0 (0.00)	8 (0.34)	89 (1.48)	147 (4.63)
0 (0.00)	1 (0.08)	11 (0.54)	18 (1.60)
0 (0.00)	11 (0.40)	230 (2.67)	459 (5.77)
36.89 (0.02)	66 (0.42)	288 (2.63)	524 (5.56)
Q1/Q2* (n=10,750)	Q3 (n=4,982)	Q4 (n=5,244)	Top 5% HCU (n=1,061)
36 (0.00)	49 (0.22)	50 (0.44)	53 (1.03)
0 (0.01)	2 (0.06)	4 (0.44)	10 (2.01)
0 (0.00)	0 (0.00)	182 (2.01)	354 (4.80)
0 (0.00)	1 (0.07)	36 (0.78)	68 (2.61)
0 (0.00)	0 (0.03)	5 (0.34)	10 (1.15)
0 (0.00)	1 (0.07)	224 (2.11)	431 (5.04)
36 (0.01)	52 (0.17)	278 (2.07)	495 (5.05)
Q1/Q2* (n=4,995)	Q3 (n=2,500)	Q4 (n=2,494)	Top 5% HCU (n=500)
38 (0.06)	58 (0.61)	52 (0.74)	57 (1.86)
1 (0.03)	4 (0.24)	10 (1.32)	31 (6.19)
0 (0.00)	11 (0.57)	220 (2.85)	374 (7.68)
0 (0.00)	32 (0.92)	74 (1.85)	144 (5.78)
0 (0.00)	1 (0.13)	7 (0.53)	12 (1.49)
0 (0.00)	45 (1.11)	301 (3.03)	530 (7.19)
39 (0.06)	108 (0.94)	364 (3.24)	619 (8.17)
Q1/Q2* (n=17,349)	Q3 (n=8,663)	Q4 (n= 8,567)	Top 5% HCU (n=1,734)
36 (0.00)	52 (0.24)	47 (0.33)	51 (0.96)
1 (0.01)	3 (0.07)	4 (0.43)	8 (2.09)
0 (0.00)	6 (0.22)	188 (1.48)	331 (3.70)
0 (0.00)	8 (0.20)	60 (0.79)	108 (2.47)
0 (0.00)	0 (0.04)	7 (0.30)	13 (0.91)
0 (0.00)	14 (0.31)	255 (1.48)	453 (3.74)
37 (0.01)	69 (0.30)	305 (1.51)	512 (3.99)

Appendix VI. Regression model

Total healthcare costs = 92.43006 + 1.979414*number of MSK diagnosis -18.14823*region none + 3.684978* region spine - 10.80206*region upper extremity - 9.017858* SES middle -1.360714*SES high + 0.827822* CDS + 6.073372* female + 0.171268* age

	Regression Coefficient (SE)	Pooled Standard Error	Pooled 95% CI	
			2.5 %	97.5 %
Intercept	92.43006	1.127535	90.22009	94.64003
Age	0.171268	0.01688466	0.1381741	0.2043619
Female	6.073372	0.5544197	4.986709	7.160035
Chronic Disease Score	0.827822	0.1122947	0.6077244	1.04792
Number of MSK diagnosis	1.979414	0.6793149	0.6479568	3.310871
Region (ref: lower extremity)				
Region none*	-18.14823	0.6575838	-19.43709	-16.85937
Region spine	3.684978	0.6749788	2.36202	5.007936
Region upper extremity	-10.80206	0.7413566	-12.25512	-9.349001
SES (ref: SES low)				
SES middle	-9.017858	0.3258425	-9.656509	-8.379207
SES high	-1.360714	0.6316623	-2.598772	-0.1226559

*This includes a musculoskeletal complaint that are not restricted to one region such as fibromyalgia, osteoporosis, and rheumatoid arthritis. R^2 = 0.007, RMSE = 127.8


3

Content Validity of Patient-Reported Outcome Measures of Satisfaction with Primary Care for Musculoskeletal Complaints: A Systematic Review

J Orthop Sports Phys Ther. 2021; doi:10.2519/jospt.2021.9788

Pellekooren S, Ostelo RWJG, Pool-Goudzwaard AL,
van Tulder MW, Jansma E, Chiarotto A.

Abstract

Background

Patient satisfaction reflects the extent to which a patient is satisfied with a treatment received. Although patient satisfaction is frequently measured in the treatment of musculoskeletal complaints satisfaction and outcomes are used to assess the quality of health care and health care providers, the best method of measuring satisfaction is unclear.

Objective

To evaluate the content validity of patient-reported outcome measurement (PROM) instruments used with patients with musculoskeletal complaints treated in primary care to assess satisfaction.

Methods

A literature search in MEDLINE, EMBASE and CINAHL was undertaken (up to January 2020) to identify studies of the development or evaluation of content validity of a PROM aimed to assess patient satisfaction. A PROM was considered eligible if it aimed to measure satisfaction with care in patients with musculoskeletal complaints. Two independent reviewers performed study selection, quality assessment, and data extraction. Evaluation of content validity of the included PROMs was performed according to COSMIN guidance, which includes the evaluation of the quality of a PROM development, the quality of content validity studies, the content of the PROMs, and rating the quality of evidence with a modified GRADE approach.

Results

Seven PROMs were identified. Their quality of development was inadequate. No studies evaluating content validity of the satisfaction PROMs were retrieved. The content validity of the satisfaction PROMs was insufficient and supported by very low-quality evidence.

Conclusions

In measuring patient satisfaction among patients with musculoskeletal complaints treated in primary care, none of the identified PROMs had adequate content validity. Future studies should address relevance, comprehensiveness, and comprehensibility of PROMs used to measure satisfaction, and emphasise patient involvement during the development of new instruments.

Introduction

Patient satisfaction reflects the extent to which a patient is satisfied with a treatment received.[1] Patient satisfaction is a multidimensional construct, defined as 'a feeling of contentment that patients experience when their perceived biopsychosocial needs have been met and their expectations of the therapeutic intervention have been fulfilled'.[2] However, many different definitions of patient satisfaction are used in research.[3]

Satisfied patients seem to gain more benefit from health care,[3] keep their appointments,[4,5] and are more adherent to treatment.[6,7] Patient satisfaction outcomes are used to assess the quality of health care and health care providers. Satisfaction outcomes play an important role in policy, which also influences health care providers' interaction with patients and health care costs.[1,8]

Patient satisfaction is frequently measured in the treatment of musculoskeletal complaints.[9,10] However, the best method of measuring satisfaction is unclear.[10–13] Measurement instruments are mostly single-item scales developed for a single study.[13,14] This may be due to the lack of an appropriate theoretical framework of the satisfaction construct from a patient's perspective. The satisfaction concept is mainly derived from consumer satisfaction and corresponding underlying theories.[2,11] As a result, the theoretical framework only partly reflects patient satisfaction in health care and may lead to misleading insights.[2,8]

Patient satisfaction is influenced by patient/provider, clinical, and contextual factors.[15] Besides issues regarding its underlying framework, there are also sources of bias that may threaten the validity of the instruments (e.g., content and format of the instruments used, sampling difficulties and social-psychological artefacts).[13] Patient satisfaction research often reports on high satisfaction levels.[9,12,16] This may be an overestimation of satisfaction due to methodological flaws and lack of a known cut-off point between satisfaction and dissatisfaction.[13]

Content validity is 'the degree to which the content of an instrument is an adequate reflection of the construct to be measured'.[17] Content validity is the first measurement property that should be assessed when selecting an instrument,[17] as it helps researchers link the content of the instrument with the construct to be measured. Although systematic reviews about patient satisfaction in musculoskeletal complaints treated in primary care have been published,[10,12] none of these studies examined the content validity of the measurement instruments and none took the

methodological quality of development and validation studies into account. We aimed to evaluate the content validity of measurement instruments used to assess satisfaction in patients with musculoskeletal complaints treated in primary care.

Methods

This review was conducted according to the COSMIN guidelines for systematic reviews of PROMs[18-21] and the COSMIN methodology for assessing the content validity of PROMs.[22] The protocol was registered in PROSPERO (registration number: CRD42019132623).

Literature Search

An initial search of MEDLINE, EMBASE and CINAHL was undertaken from time of inception of the database until May 2019. We updated the search in January 2020. Free text words and index terms for patient satisfaction, musculoskeletal disorders, and primary care were combined with an adjusted PROM filter[23] and a construct validity filter based on the filter of Terwee et al[24] to identify studies on content validity. The search was conducted under supervision of an experienced clinical librarian (EJ). Extensive details of the search are presented in Appendix I. The reference lists of all included articles were screened for additional relevant studies. We used Google Scholar for forward citation tracking to retrieve possibly relevant studies that were recently published. Results of all searches were combined and de-duplicated in EndNote X8.2.

Eligibility Criteria

Studies that concerned the development or the evaluation of the content validity of a PROM (in terms of relevance, comprehensiveness, and comprehensibility) that aim to assess patient satisfaction were considered for inclusion. Patient satisfaction refers to both clinical outcome and contextual factors[15] that influence the feeling of contentment that patients experience when their perceived biopsychosocial needs have been met and their expectations of the therapeutic intervention have been fulfilled'.[2] The study population had to include adult patients (≥ 18 years old) with musculoskeletal complaints (e.g. low back pain, shoulder- and knee complaints) treated in a primary care setting (e.g. general practice, physiotherapy, manual therapy, chiropractic). Studies had to be published as full-text article in English or Dutch. Studies that concerned translation or cross-cultural adaptation of a PROM were also included if a pilot study to evaluate the comprehensibility of the cross-culturally adapted PROM was performed.

Article Selection

A first decision on eligibility was based on title and abstract. For studies meeting the inclusion criteria the full-text manuscript was obtained. If a decision could not be made based on the title and abstract, the full text paper was obtained to make a decision

on eligibility. A standardized selection form was used. Studies were distributed evenly among the pairs of reviewers (SP/ RO and MT / AP) and selected independently. Disagreements regarding inclusion were discussed and resolved by consensus; if necessary, a third reviewer (AC) was consulted to make a final decision. We recorded the reasons for exclusion of full-texts, and the studies included. Results of the selection process were summarised in a flow chart.[25] References were managed in EndNote X8.2.

Data Extraction

A standardised data extraction form[22] was used. In order to achieve a uniform method of data extraction, a pilot was undertaken using a related but not eligible study. The following information was extracted: characteristics of the PROM (i.e., construct, type of measurement model (formative or reflective), target population, intended context of use, number of scales, number of items and recall period) and characteristics of the study (i.e., study population, sample size, study design, patients and/or professionals involved). In formative models, items form the construct, and changing the items will change the meaning or interpretation of the construct. In reflective models, the items are a reflection of the construct, and changing the items will not alter the meaning or interpretation of the construct.[26] In attributing the type of model, the reviewers (SP, AC) assessed if changing the construct influenced the items. If so, the model was seen as reflective. When changing the construct did not influence all items the model was seen as formative. SP performed the initial assessment of type of model (i.e., reflective or formative) by screening all items of each PROM. AC doublechecked all the PROM characteristics, including type of model. A decision was made on consensus, which was the case in all PROMS.

Evaluation of Content Validity of the Included PROM

During the evaluation of content validity according to the COSMIN guidelines,[22] the following three steps were performed: 1. evaluating the quality of the PROM development; 2. evaluating the quality of additional content validity studies of the PROM; and 3. evaluating the content validity of the PROM, based on the quality and results of the available studies and the PROM itself. In order to achieve a uniform method of evaluating content validity, a pilot was undertaken by using a related non-included study. During the evaluation of the content validity, two researchers independently (SP, AC) rated the content validity of each PROM. Disagreements between assessors were discussed and resolved by consensus. If necessary, a third reviewer (RO) was consulted.

Step 1; Evaluate the Quality of the PROM Development

First, the COSMIN website (https://database.cosmin.nl) was checked to establish whether the quality of the PROM development was already rated by others. If rating of the PROM development already existed, this rating was used instead of rating the

quality of the PROM development ourselves. If a rating was absent, box 1 of the COSMIN guidelines[22] was used for rating the quality of the PROM development study. During the rating process the following two aspects were evaluated: a. The concept elicitation, i.e., quality of research performed to identify relevant items for a new PROM and b. The cognitive interview study, i.e., quality of interviews or other pilot test (e.g., survey) performed to evaluate comprehensiveness and comprehensibility of the PROM. These findings were included in the final rating of the quality of the PROM development.

Step 2; Evaluating the Quality of Additional Content Validity Studies on the PROM
Depending on the available information found in the content validity studies, a decision was made on which of the five parts of box 2 of the COSMIN guidelines[22] had to be completed. The evaluation concerned the following five parts: a. asking patients about the relevance of the PROM items, b. asking patients about the comprehensiveness of the PROM, c. asking patient about the comprehensibility of the PROM, d. asking professionals about the relevance, and e. asking professionals about the comprehensiveness of the PROM.

Step 3; Evaluating the Content Validity of the PROM
Based on a summary of all available evidence of the PROM development (i.e., Step 1) and additional content validity studies (i.e., Step 2), the content validity of the PROM was rated. Each score or sub score of a PROM was rated separately. The rating of the content validity of a PROM contained three sub-steps. First, each result of a single study on PROM development and content validity was rated against the 10 criteria for good content validity.[22] These criteria concerned five criteria on relevance (i.e.; are the included items relevant for the construct of interest, are the included items relevant for the target population of interest, are the included items relevant for the context of use, are the response options appropriate, and is the recall period appropriate), one criterion on comprehensiveness (i.e.; are no key concepts missing), and four criteria on comprehensibility (i.e.; are the PROM instructions understood by the population of interest as intended, are the PROM items and response options understood by the population of interest as intended, are the PROM items appropriately worded, and do the response options match the question).

In rating the relevance of the items, we used six themes:[15]

a. outcomes (e.g., items referring to recovery and pain relief)
b. physiotherapist features (e.g., items referring to human competence, such as being respectful, and items referring to professional competence, such as having knowledge of the most effective treatment)
c. patient features (e.g., items referring to meeting patient expectations)
d. caregiver-patient relationship (e.g., items referring to communication skills and partnership of care)

e. treatment features (e.g., items referring to patient education an organisation of care)
f. healthcare setting features (e.g., items referring to the physical environment and social context).

This assessment was conducted by two independent reviewers (SP, AC). Subsequently, the reviewers scored the relevance, comprehensiveness, and comprehensibility of the PROMs independently for content validity (10 items),[22] and the six relevance themes.[15] Second, results of all studies were qualitatively summarised to determine whether the overall relevance, overall comprehensiveness, overall comprehensibility, and overall content validity was sufficient (+), insufficient (–), inconsistent (±), or indeterminate (?), taking all evidence into account. Last, overall ratings determined in previous steps were graded for the quality of the evidence using a modified GRADE approach (i.e., high, moderate, low, very low-quality evidence).

Results

Literature Search and Selection

After screening the titles and abstracts of 5772 articles, 40 studies were retrieved for full-text assessment. Thirty-four studies did not meet the inclusion criteria. After reference checking and citation searching one more study was included. Eventually, seven studies[27–33] of seven different PROMs were eligible for assessment of content validity. Details of the search are presented in Figure 1.

Characteristics of the Included Studies

Six studies focused on patient satisfaction with physiotherapy treatment [27,29–33] and one study focused on chiropractic treatment.[28] All but one study,[31] based their instrument on a formative model, supplemented with reflective measures. More extensive details on the PROM characteristics are presented in Appendix II, and details on characteristics of the included study populations are presented in Table 1.

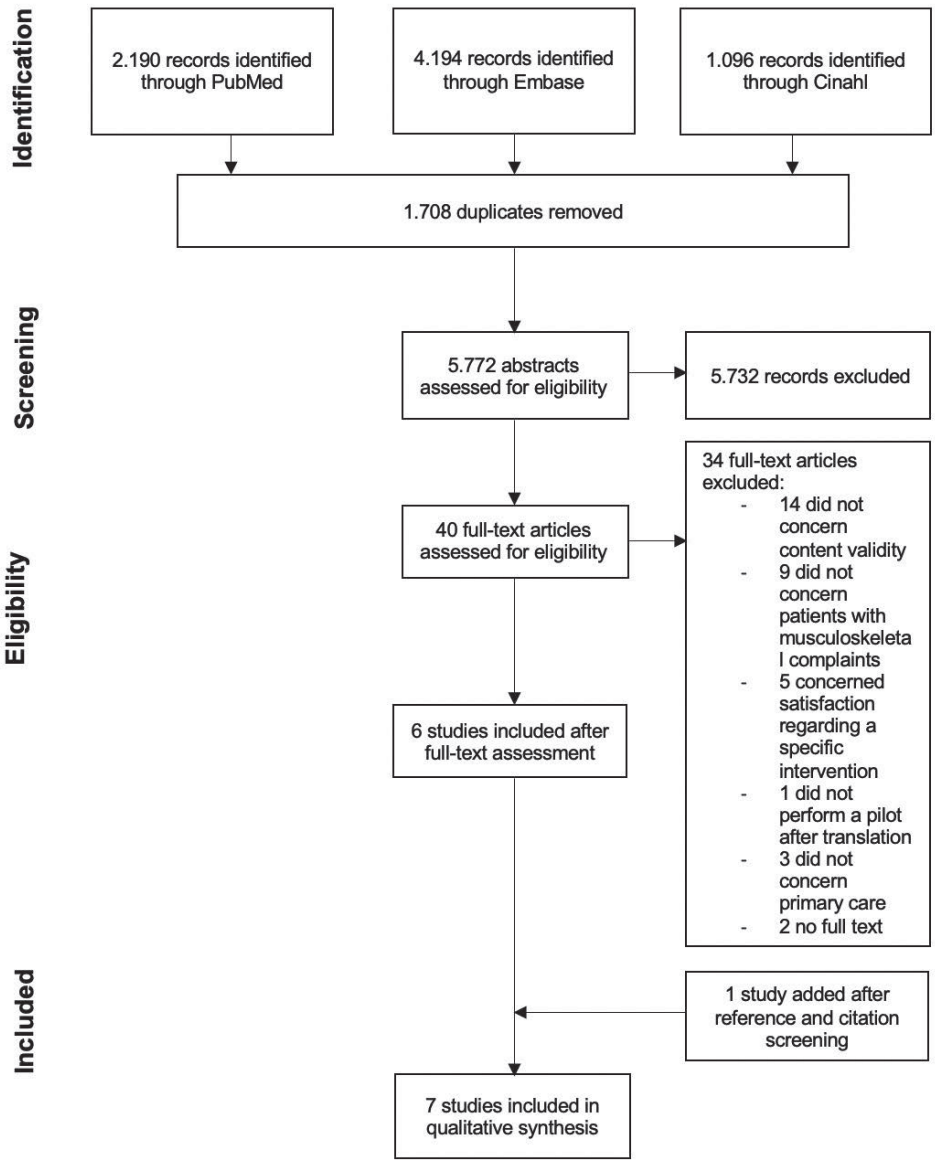


Figure 1. Flowchart of studies includes

Quality and Results of Content Validity Studies

No content validity studies on existing patient satisfaction PROMs were identified.

Rating PROMs Against the Criteria of Content Validity

All of the included PROMs [27–33] were judged 'indeterminate' for relevance, as there was no clear description of the construct, no clarity on the origin of the construct, and there was no evidence from concept elicitation, literature, or professionals that at least 85% of the items of the instrument refer to the construct patient satisfaction. None of the PROMs provided a clearly described and representative target population, which was involved in the elicitation of relevant items. There was no clear description of the context of use for any of the PROMs, and there was no justification for the response options or recall period. Patients were not asked about the comprehensiveness of the PROM in an adequate cognitive interview and more than 15% of the key concepts were missing, which resulted in an 'indeterminate' rating on comprehensiveness for all studies. We judged all PROMs as 'indeterminate' for comprehensibility, as during the development of these PROMs patients were not asked about comprehensibility of the instructions, items, and response options in an adequate cognitive interview.

In the part of the reviewers' rating, we judged that fewer than 85% of the PROMs items were relevant for the construct, population, and context of use. At least 85% of the response options were appropriate for all PROMs, as they were in accordance with the questions and provided a reasonable range of response options. None of the PROMs provided a justification for the recall period, or the lack of it. All of the included PROMs were 'insufficient' for relevance. Regarding comprehensiveness, none of the PROMs contained more than 85% of the key concepts identified by Rossettini et al,[15] and we judged all PROMs 'insufficient' for comprehensiveness. All PROMs had 'sufficient' comprehensibility, as we considered the response options appropriately worded (e.g., reading level, ambiguous terms, and jargon), and at least 85% of the response options matched the questions. More extensive details on the criteria of content validity of the PROM are presented in Table 3.

Evidence Synthesis

The quality of evidence was 'very low' for all PROMs, as all of the development studies were rated as 'inadequate' and were not complemented with additional studies on their content validity.[22] Overall content validity was 'insufficient' for all PROMs. More extensive details on the modified GRADE approach are presented in Table 3.

Table 1. Characteristics of the included study populations

PROM	Ref	Population		
		N	Age Mean (SD) year	Gender % Female
Patient Satisfaction with Outpatient Physical Therapy Instrument	Beatie 2002[27]	Pilot study: 191 Validation study: 1,868	Pilot study: NA Validation study: 46.9 (11.9)	Pilot study: NA Validation study: 36,2 %
MedRisk Instrument to Measure Patient Satisfaction with Chiropractic Care (MRPS-CC)	Beattie 2011[28]	323	Female 47.2 (11.6) Male 49.7 (10.7)	50.2%
Physical Therapy Patient Satisfaction Questionnaire	Goldstein 2000[29]	289	45.7 (17.3)	63.7%
Patient satisfaction with physical therapy instrument	Mendonca 2007[30]	Pilot study: 175 Validation study: 834	Pilot study; NA Validation study: 46,7(15,8)	Pilot study: NA Validation study: 64,4%

Disease characteristics			Instrument administration			
Disease	**Disease duration mean (SD) yr**	**Disease severity**	**Setting**	**Country**	**Language**	**Response rate**
Pilot study: Musculoskeletal conditions Validation study: Musculoskeletal conditions	NA	NA	Outpatient physical therapy clinics	United States	English	20%
Musculoskeletal conditions Area that was treated: 45% Back 42% Neck 11% Upper extremity 2% Spine and extremity	NA	NA	Chiropractic clinics	United States	English	NA
Musculoskeletal conditions Area that was treated: 31.1% Back 26.6 % Shoulder 19.4% Knee 18.7% Neck	NA	NA	Outpatient physical therapy clinics	United States	English	NA
Pilot study: NA Validation study: 57.4% Orthopedics/ traumatology 22.5% Rheumatology 7.0% Neurology	NA	NA	Outpatient physical therapy clinics	Brazil	Portuguese	Validation study; 60%

Table 1. Continued

PROM	Ref	Population		
		N	Age Mean (SD) year	Gender % Female
Scale to measure patient satisfaction with physical therapy	Monnin 2002[31]	524* * of 501 patients' demographic information was available	58.6 (18.9)	48.5%
Physical Therapy Outpatient Satisfaction Survey (PTOPS)	Roush 1999[32]	Phase 1: 177 Phase 2: 257 Phase 3: 63	Phase 1: 45.2 (15.6) Phase 2: 46.7 (16.3) Phase 3: 49.4 (16.9)	Phase 1: 60% Phase 2: 57% Phase 3: 64%
Satisfaction with physiotherapy Questionnaire	Sadeq 2002[33]	144	40 (12)	64%

NA: information is not available.

Disease characteristics			Instrument administration			
Disease	**Disease duration mean (SD) yr**	**Disease severity**	**Setting**	**Country**	**Language**	**Response rate**
58.2% Outpatients 41.8% Inpatients Type of rehabilitation: 36.8% medical 29.8% orthopaedic 17.9% cardiorespiratory 15.5% neurological	NA	NA	Out and in patients University Hospital	Switzerland	French	52%
Musculoskeletal conditions	NA	NA	Outpatient physical therapy clinics	United States	English	Phase 1: 69% Phase 2: 80% Phase 3: 76%
Musculoskeletal conditions	NA	NA	Outpatient physiotherapy department University Hospital	Kuwait	English and Arabic	NA

Table 2. Quality of the PROM development

PROM
Patient Satisfaction with Outpatient Physical Therapy Instrument[27]
MedRisk Instrument to Measure Patient Satisfaction with Chiropractic Care (MRPS-CC)[28]
Physical Therapy Patient Satisfaction Questionnaire[29]
Patient satisfaction with physical therapy instrument[30]
Scale to measure patient satisfaction with physical therapy[31]
Physical Therapy Outpatient Satisfaction Survey (PTOPS)[32]
Satisfaction with physiotherapy Questionnaire[33]

Abbreviations: VG, very good; A, adequate; D, doubtful; I, inadequate; NA, not applicable.
* When the PROM was not developed in a sample representing the target population, the concept elicitation was not further rated

PROM design							Total PROM development
General design requirements					Concept elicitation*	Total PROM design	
Clear construct	Clear origin of construct	Clear target population	Clear context of use	Sample representing the target population			
I	D	VG	D	A	I	I	I
I	D	VG	D	I		I	I
VG	VG	VG	VG	I		I	I
I	D	VG	VG	I		I	I
I	D	I	D	I		I	I
I	D	I	D	I		I	I
I	D	I	D	A	I	I	I

Table 3. Ten criteria of content validity and grading the quality of the evidence

PROM
Patient Satisfaction with Outpatient Physical Therapy Instrument[27]
MedRisk Instrument to Measure Patient Satisfaction with Chiropractic Care(MRPS-CC)[28]
Physical Therapy Patient Satisfaction Questionnaire[29]
Patient satisfaction with physical therapy instrument[30]
Scale to measure patient satisfaction with physical therapy[31]
Physical Therapy Outpatient Satisfaction Survey (PTOPS)[32]
Satisfaction with physiotherapy Questionnaire[33]

Abbreviations: +, sufficient; -, insufficient; ?, indeterminate; ±, inconsistent.
* Quality of evidence:
High quality; At least one content validity study of very good or adequate quality
Moderate quality; At least one content study of doubtful quality OR only content studies of inadequate quality or no content studies and PROM development study of very good or adequate quality

During the assessment on PROM development and criteria on content validity there was agreement on 107 items, and disagreement on 24 items (78% absolute agreement). In rating the ten criteria on content validity, 56 items were not further assessed, as the COSMIN guidelines recommend to rate PROM development studies of inadequate quality as 'indeterminate'. Disagreements were discussed and resolved by consensus between the two reviewers. There was no need to consult a third reviewer during the assessment of item.

PROM development study			Rating of reviewers			Overall rating content validity	Quality of evidence*
Relevance	Comprehensiveness	Comprehensibility	Relevance	Comprehensiveness	Comprehensibility		
?	?	?	-	-	+	-	very low
?	?	?	-	-	+	-	very low
?	?	?	-	-	+	-	very low
?	?	?	-	-	+	-	very low
?	?	?	-	-	+	-	very low
?	?	?	-	-	+	-	very low
?	?	?	-	-	+	-	very low

Low quality: Only content validity studies of inadequate quality OR no content validity studies and PROM development study of doubtful quality

Very low quality: Only content validity studies of inadequate quality OR no content validity studies and PROM development study of inadequate quality

Discussion

We aimed to evaluate the content validity of instruments that assess satisfaction of patients with musculoskeletal complaints treated in primary care. Although patient satisfaction seems to be an important goal in healthcare, we only identified seven PROMs, and no PROM had sufficient content validity. The quality of development studies of these PROMs was 'inadequate', the content of all PROMs was 'insufficient', and these ratings were supported by 'very low' quality evidence.

Shortcomings concerned various aspects regarding the relevance, comprehensiveness, and comprehensibility of PROMs. None of the studies in our systematic review included patients during the whole process of developing the PROM. Some developers argue that patients find it difficult to assess quality of healthcare, and therefore possibly overemphasize some elements and dismiss others.[13] However, empirical evidence indicates that patients may be more capable judges than some may imply.[13] If the construct aims to measure patient satisfaction, patients are the key stakeholders in judging whether they are satisfied with the treatment. Ignoring patients' opinions, while aiming to capture their experiences, seems contradictory. None of the PROMs included items that were related to 'shared decision making', which is a key issue in current health care.

Most PROMs were based on a formative model, which assumes that all items included fully form the framework patient satisfaction. One may argue that a reflective model is more appropriate in a complex theoretical framework, such as patient satisfaction, because reflective models take unexplained variance into account, in contrast to formative models. [26,34]

In all PROMs, satisfaction was assumed to be a continuous variable. The point at which satisfaction stops and dissatisfaction begins is unclear,[13] and interpretation of PROM scores is difficult. Additionally, given the lack of gold standards for measuring satisfaction, using arbitrary cut-off scores between 'satisfied' and 'not satisfied' could lead to an overestimation of the level of patient satisfaction.[12] We encourage researchers to use analyse satisfaction PROMs as continuous outcomes, although further psychometric analyses of these PROMs (e.g., Rasch analysis) is required to clearly demonstrate that these scores can be used as such.

The quality of evidence was influenced by the lack of additional content validity studies. Validating the developed PROMs may be difficult and sometimes obsolete. Older PROMs may not comply with current perspectives on health care as the concept 'quality of care'

has changed considerably during last few decades. Additionally, researchers may be less motivated to conduct ex-novo studies on PROMs that were developed for a more generic population in a determined clinical setting, and which may not be available in other languages.

Although there are numerous studies published on patient satisfaction in musculoskeletal healthcare,[12] none of these studies report on content validity or other clinimetrical properties of a PROM. Only a few qualitative studies[15,35,36] on patient satisfaction have evaluated possible determinants of patient satisfaction. These studies clearly are of great value in the development of items, as their findings are based on patient opinions.

Limitations

As there is no clear definition of the construct patient satisfaction, we used items from Rosettini et al[15] to rate relevance of items. Although Rosettini's review provides a thorough overview of factors influencing patient satisfaction, it may not have identified all factors that contribute to the construct patient satisfaction. It is possible we did not rate all relevant factors to content validity of the PROMs. As measurement of patient satisfaction is often part of a broader quality evaluation within healthcare centres, not all PROMs used to measure patient satisfaction are published. Also, not all PROMs for measuring patient satisfaction are freely accessible. For these reasons, it is very likely that a number of PROMs were not included in this study, and that our findings only represent a small part of all available PROMs measuring patient satisfaction in patients with musculoskeletal complaints treated in primary care. Nevertheless, the added value of including these PROMs in this review may be relative, as there may be a lack of content validity studies regarding these PROMs.

Implications

We alert clinicians that any PROM measuring the satisfaction of patients with musculoskeletal disorders should be used with caution, as it is unclear if these instruments comprehensively reflect the satisfaction construct. We also encourage clinicians to cautiously interpret patient satisfaction scores. Given the low content validity of these PROMs, we encourage clinicians to augment satisfaction evaluation by asking patients in more detail if they are satisfied with the treatment received, and why. This might be a temporary alternative until a fit-for-purpose instrument is available.

Future research could benefit from using the COSMIN guidelines[22] in the development of a PROM. These guidelines emphasize the importance of patient involvement during the whole development and clarifies the different aspects of content validity of a PROM

(i.e., relevance, comprehensiveness, and comprehensibility). In addition, one could consider using a reflective model to account for unexplained variance. Validation of PROMs would be highly desirable, as this contributes to strength of evidence.

Conclusion

Seven PROMs used to measure patient satisfaction had 'insufficient' content validity and were supported by 'very low' quality of evidence due to shortcomings in the development of the PROMs as well as the lack of validation studies. Clinicians should use these PROMs with caution and be careful in interpreting their scores. There is a dire need for adequately developing and validating PROMs to assess satisfaction in patients with musculoskeletal disorders.

References

1. Bachman J. The Problem with Patient Satisfaction Scores. *Fam Pract Manag*. 2016;23:23-27.
2. Hills R, Kitchen S. Toward a theory of patient satisfaction with physiotherapy: Exploring the concept of satisfaction. *Physiother Theory Pract*. 2007;23(5):243-254. doi:10.1080/09593980701209394
3. Fitzpatrick R. The assessment of patient satisfaction. In: Jenkinson J, ed. Assessment and Evaluation of Health and Medical Care. Open University Press; 1997:85-102.
4. Bigby J. Medical Consequences of Missed Appointments. *Arch Intern Med*. 1984;144(6):1163. doi:10.1001/archinte.1984.00350180075012
5. Frankel S, Farrow A, West R. Non-attendance or non-invitation? A case-control study of failed outpatient appointments. *BMJ*. 1989;298(6684):1343-1345. doi:10.1136/bmj.298.6684.1343
6. Kincey J, Bradshaw P, Ley P. Patients' satisfaction and reported acceptance of advice in general practice. *J R Coll Gen Pract*. 1975;25:558-566.
7. Willson P, McNamara JR. How perceptions of a simulated physician-patient interaction influence intended satisfaction and compliance. *Soc Sci Med*. 1982;16(19):1699-1704. doi:10.1016/0277-9536(82)90095-8
8. Kupfer JM, Bond EU. Patient Satisfaction and Patient-Centered Care. *JAMA*. 2012;308(2):139. doi:10.1001/jama.2012.7381
9. Desmeules F, Roy JS, MacDermid JC, Champagne F, Hinse O, Woodhouse LJ. Advanced practice physiotherapy in patients with musculoskeletal disorders: a systematic review. *BMC Musculoskelet Disord*. 2012;13(1):107. doi:10.1186/1471-2474-13-107
10. Fennelly O, Blake C, Desmeules F, Stokes D, Cunningham C. Patient-reported outcome measures in advanced musculoskeletal physiotherapy practice: a systematic review. *Musculoskeletal Care*. 2018;16(1):188-208. doi:10.1002/msc.1200
11. Batbaatar E, Dorjdagva J, Luvsannyam A, Amenta P. Conceptualisation of patient satisfaction: a systematic narrative literature review. *Perspect Public Health*. 2015;135(5):243-250. doi:10.1177/1757913915594196
12. Hush JM, Cameron K, Mackey M. Patient Satisfaction With Musculoskeletal Physical Therapy Care: A Systematic Review. *Phys Ther*. 2011;91(1):25-36. doi:10.2522/ptj.20100061
13. Pascoe GC. Patient satisfaction in primary health care: A literature review and analysis. *Eval Program Plann*. 1983;6(3-4):185-210. doi:10.1016/0149-7189(83)90002-2
14. Crow R, Gage H, Hampson S, et al. The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature. *Health Technol Assess (Rockv)*. 2002;6(32). doi:10.3310/hta6320
15. Rossettini G, Latini TM, Palese A, et al. Determinants of patient satisfaction in outpatient musculoskeletal physiotherapy: a systematic, qualitative meta-summary, and meta-synthesis. *Disabil Rehabil*. 2020;42(4):460-472. doi:10.1080/09638288.2018.1501102
16. Marks D, Comans T, Bisset L, Scuffham PA. Substitution of doctors with physiotherapists in the management of common musculoskeletal disorders: a systematic review. *Physiotherapy*. 2017;103(4):341-351. doi:10.1016/j.physio.2016.11.006
17. Prinsen CAC, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" – a practical guideline. *Trials*. 2016;17(1):449. doi:10.1186/s13063-016-1555-2
18. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*. 2018;27(5):1147-1157. doi:10.1007/s11136-018-1798-3

19. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research*. 2018;27(5):1171-1179. doi:10.1007/s11136-017-1765-4
20. Mokkink LB, Terwee CB, Patrick 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;63(7):737-745. doi:10.1016/j.jclinepi.2010.02.006
21. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research*. 2018;27(5):1159-1170. doi:10.1007/s11136-018-1829-0
22. Terwee C, Prinsen C, Chiarotto A, et al. COSMIN methodology for assessing the content validity of PROMs. Accessed January 11, 2020. https://www.cosmin.nl/tools/guideline-conducting-systematic-review-outcome-measures/
23. Mackintosh A, Casañas i Comabella C, Hadi M, Gibbons E, Fitzpatrick R. Patient-Reported Outcome Measurement. Oxford University Group Department of Public Health. Published 2010. Accessed November 1, 2020. https://cosmin.nl/wp-content/uploads/prom-search-filter-oxford-2010.pdf
24. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research*. 2009;18(8):1115-1123. doi:10.1007/s11136-009-9528-5
25. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *J Clin Epidemiol*. 2009;62(10):1006-1012. doi:10.1016/j.jclinepi.2009.06.005
26. Edwards JR. The Fallacy of Formative Measurement. *Organ Res Methods*. 2011;14(2):370-388. doi:10.1177/1094428110378369
27. Beattie PF, Pinto MB, Nelson MK, Nelson R. Patient Satisfaction With Outpatient Physical Therapy: Instrument Validation. *Phys Ther*. 2002;82(6):557-565. doi:10.1093/ptj/82.6.557
28. Beattie PF, Nelson R, Murphy DR. Development and Preliminary Validation of the MedRisk Instrument to Measure Patient Satisfaction with Chiropractic Care. *J Manipulative Physiol Ther*. 2011;34(1):23-29. doi:10.1016/j.jmpt.2010.09.003
29. Goldstein MS, Elliott SD, Guccione AA. The Development of an Instrument to Measure Satisfaction With Physical Therapy. *Phys Ther*. 2000;80(9):853-863. doi:10.1093/ptj/80.9.853
30. Mendonça K, Guerra R. Desenvolvimento e validação de um instrumento de medida da satisfação do paciente com a fisioterapia. *Revista Brasileira de Fisioterapia*. 2007;11(5). doi:10.1590/S1413-35552007000500007
31. Monnin D, Perneger T V. Scale to Measure Patient Satisfaction With Physical Therapy. *Phys Ther*. 2002;82(7):682-691. doi:10.1093/ptj/82.7.682
32. Roush SE, Sonstroem RJ. Development of the Physical Therapy Outpatient Satisfaction Survey (PTOPS). *Phys Ther*. 1999;79(2):159-170. doi:10.1093/ptj/79.2.159
33. Sadeq M, Adib S. Satisfaction with physiotherapy among patients at a general hospital in Kuwait. *Saudi Med J*. 2002;23:891-985.
34. Howell RD, Breivik E, Wilcox JB. Reconsidering formative measurement. *Psychol Methods*. 2007;12(2):205-218. doi:10.1037/1082-989X.12.2.205
35. Knight PK, Cheng ANJ, Lee GM. Results of a survey of client satisfaction with outpatient physiotherapy care. *Physiother Theory Pract*. 2010;26(5):297-307. doi:10.3109/09593980903164058
36. Hills R. Patient Satisfaction with Outpatient Physiotherapy: An Examination of the Needs and Expectations of Patients with Acute and Chronic Musculoskeletal Conditions. University of London; 2003.

Appendix I. Search details

PubMed

#1 Musculoskeletal caregivers

("Physical Therapy Specialty"[MeSH] OR "Physical Therapy Modalities"[MeSH] OR "Musculoskeletal Manipulations"[MeSH]) OR Chiropractic[MeSH] OR physiotherap*[tiab] OR ((physical[tiab] OR Musculoskeletal[tiab] OR manipulati*[tiab] OR manual[tiab]) AND (therapy[tiab] OR therapies[tiab] OR activity[tiab] OR activities[tiab] OR exercis*[tiab] OR training*[tiab])) OR Musculoskeletal Manipulation*[tiab] OR physician exten*[tiab] OR specialist practitioner*[tiab] OR chiropractic*[tiab])

#2 PROM

"Surveys and Questionnaires"[Mesh:NoExp] OR Questionnaire*[tiab] OR Survey*[tiab] OR Patient Reported Outcome*[tiab] OR ((patient[tiab] OR self[tiab] OR carer[tiab] OR proxy[tiab]) AND (report[tiab] OR reported[tiab] OR reporting[tiab] OR rated[tiab] OR rating[tiab] OR ratings[tiab] OR based[tiab] OR assessed[tiab] OR assessment[tiab] OR assessments[tiab]) AND (index[tiab] OR indices[tiab] OR instrument[tiab] OR instruments[tiab] OR measure[tiab] OR measures[tiab] OR scale[tiab] OR scales[tiab] OR score[tiab] OR scores[tiab] OR status[tiab]))

#3 Patient satisfaction

"Patient-Centered Care"[Mesh] OR "Patient Satisfaction"[MeSH] OR patient-centered[tiab] OR patient satisfaction[tiab] OR patient preference*[tiab]

#4 Construct validity filter

"Validation Studies"[pt] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR Validation [tiab] OR (construct[tiab] AND valid*[tiab]) OR "outcome assessment (health care)"[MeSH] OR "outcome assessment"[tiab] OR outcome measure*[tw] OR Quality Indicators, Health Care/standards[MeSH] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tw] OR precision[tw] OR imprecision[tw] OR "precise values"[tw] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor analyses"[tiab] OR "factor structure"[tiab] OR "factor structures"[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR "item discriminant"[tiab] OR "interscale correlation*"[tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR "interval variability"[tiab] OR "rate variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR "item response model"[tiab] OR

irt[tiab] OR rasch[tiab] OR "differential item functioning"[tiab] OR dif[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab] OR (item*[ti] AND development*[ti])

NOT ("addresses"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "directory"[Publication Type] OR "editorial"[Publication Type] OR "festschrift"[Publication Type] OR "interview"[Publication Type] OR "lectures"[Publication Type] OR "legal cases"[Publication Type] OR "legislation"[Publication Type] OR "letter"[Publication Type] OR "news"[Publication Type] OR "newspaper article"[Publication Type] OR "patient education handout"[Publication Type] OR "popular works"[Publication Type] OR "congresses"[Publication Type] OR "consensus development conference"[Publication Type] OR "consensus development conference, nih"[Publication Type] OR "practice guideline"[Publication Type]) NOT ("animals"[MeSH Terms] NOT "humans"[MeSH Terms]))

EMBASE

#1 Musculoskeletal caregivers

'physiotherapy'/exp OR 'musculoskeletal manipulation'/exp OR 'chiropractic'/exp OR physiotherap*:ti,ab,kw OR ((physical:ti,ab,kw OR Musculoskeletal:ti,ab,kw OR manipulati*:ti,ab,kw OR manual:ti,ab,kw) AND (therapy:ti,ab,kw OR therapies:ti,ab,kw OR activity:ti,ab,kw OR activities:ti,ab,kw OR exercis*:ti,ab,kw OR training*:ti,ab,kw)) OR 'Musculoskeletal Manipulation*':ti,ab,kw OR 'physician exten*':ti,ab,kw OR 'specialist practitioner*':ti,ab,kw OR chiropractic*:ti,ab,kw

#2 PROM

'questionnaire'/exp OR Questionnaire*:ti,ab,kw OR Survey*:ti,ab,kw OR 'Patient Reported Outcome*':ti,ab,kw OR ((patient:ti,ab,kw OR self:ti,ab,kw OR carer:ti,ab,kw OR proxy:ti,ab,kw) AND (report:ti,ab,kw OR reported:ti,ab,kw OR reporting:ti,ab,kw OR rated:ti,ab,kw OR rating:ti,ab,kw OR ratings:ti,ab,kw OR based:ti,ab,kw OR assessed:ti,ab,kw OR assessment:ti,ab,kw OR assessments:ti,ab,kw) AND (index:ti,ab,kw OR indices:ti,ab,kw OR instrument:ti,ab,kw OR instruments:ti,ab,kw OR measure:ti,ab,kw OR measures:ti,ab,kw OR scale:ti,ab,kw OR scales:ti,ab,kw OR score:ti,ab,kw OR scores:ti,ab,kw OR status:ti,ab,kw))

#3 Patient satisfaction

'patient care'/de OR 'patient assessment'/exp OR 'patient decision making'/exp OR 'patient satisfaction'/exp OR 'patient-centered':ti,ab,kw OR 'patient satisfaction':ti,ab,kw OR 'patient preference*':ti,ab,kw

#4 Construct validity filter
'validation study'/exp OR 'psychometry'/exp OR psychometr*:ti,ab,kw OR clinimetr*:ti,ab,kw OR clinometr*:ti,ab,kw OR Validation:ti,ab,kw OR (construct:ti,ab,kw AND valid*:ti,ab,kw) OR 'outcome assessment'/exp OR 'outcome assessment':ti,ab,kw OR 'outcome measure*':ti,ab,kw OR 'health care quality'/exp OR 'internal consistency':ti,ab,kw OR (cronbach*:ti,ab,kw AND (alpha:ti,ab,kw OR alphas:ti,ab,kw)) OR (item:ti,ab,kw AND (correlation*:ti,ab,kw OR selection*:ti,ab,kw OR reduction*:ti,ab,kw)) OR agreement:ti,ab,kw OR precision:ti,ab,kw OR imprecision:ti,ab,kw OR 'precise values':ti,ab,kw OR (intraclass:ti,ab,kw AND correlation*:ti,ab,kw) OR discriminative:ti,ab,kw OR 'known group':ti,ab,kw OR 'factor analysis':ti,ab,kw OR 'factor analyses':ti,ab,kw OR 'factor structure':ti,ab,kw OR 'factor structures':ti,ab,kw OR dimension*:ti,ab,kw OR subscale*:ti,ab,kw OR (multitrait:ti,ab,kw AND scaling:ti,ab,kw AND (analysis:ti,ab,kw OR analyses:ti,ab,kw)) OR 'item discriminant':ti,ab,kw OR 'interscale correlation*':ti,ab,kw OR error:ti,ab,kw OR errors:ti,ab,kw OR 'individual variability':ti,ab,kw OR 'interval variability':ti,ab,kw OR 'rate variability':ti,ab,kw OR (variability:ti,ab,kw AND (analysis:ti,ab,kw OR values:ti,ab,kw)) OR 'item response model':ti,ab,kw OR irt:ti,ab,kw OR rasch:ti,ab,kw OR 'differential item functioning':ti,ab,kw OR dif:ti,ab,kw OR 'computer adaptive testing':ti,ab,kw OR 'item bank':ti,ab,kw OR 'cross-cultural equivalence':ti,ab,kw OR (item*:ti AND development*:ti)

Cinahl

#1 Musculoskeletal caregivers
(MH "Physical Therapy") OR (MH "Physical Therapy Practice, Research-Based") OR (MH "Physical Therapy Assessment") OR (MH "Research, Physical Therapy") OR (MH "Physical Therapy Practice") OR (MH "Physical Therapy Service") OR (MH "Chiropractic") OR (MH "Manipulation, Chiropractic") OR (MH "Research, Chiropractic") OR (MH "Chiropractic Practice") OR (MH "Chiropractors") OR TI (physiotherap*) OR (TI ("physical" OR "Musculoskeletal" OR "manipulati*" OR "manual") AND TI ("therapy" OR "therapies" OR "activity" OR "activities" OR "exercis*" OR "training*")) OR TI ("Musculoskeletal Manipulation*" OR "physician exten*" OR "specialist practitioner*" OR "chiropractic*") OR AB (physiotherap*) OR (AB ("physical" OR "Musculoskeletal" OR "manipulati*" OR "manual") AND AB ("therapy" OR "therapies" OR "activity" OR "activities" OR "exercis*" OR "training*")) OR AB ("Musculoskeletal Manipulation*" OR "physician exten*" OR "specialist practitioner*" OR "chiropractic*")

#2 PROM
(MH "Surveys") OR (MH "Questionnaires+") OR TI ("Questionnaire*" OR "Survey*" OR "Patient Reported Outcome*" OR (TI ("patient" OR "self" OR "carer" OR "proxy") AND TI ("report" OR "reported" OR "reporting" OR "rated" OR "rating" OR "ratings" OR "based"

OR "assessed" OR "assessment" OR "assessments") AND TI ("index" OR "indices" OR "instrument" OR "instruments" OR "measure" OR "measures" OR "scale" OR "scales" OR "score" OR "scores" OR "status")) OR AB ("Questionnaire*" OR "Survey*" OR "Patient Reported Outcome*" OR (AB ("patient" OR "self" OR "carer" OR "proxy") AND AB ("report" OR "reported" OR "reporting" OR "rated" OR "rating" OR "ratings" OR "based" OR "assessed" OR "assessment" OR "assessments") AND AB ("index" OR "indices" OR "instrument" OR "instruments" OR "measure" OR "measures" OR "scale" OR "scales" OR "score" OR "scores" OR "status"))

#3 Patient satisfaction
(MH "Patient Centered Care") OR (MH "Patient Satisfaction+") OR TI ("patient-centered" OR "patient satisfaction" OR "patient preference*") OR AB ("patient-centered" OR "patient satisfaction" OR "patient preference*")

#4 Construct validity filter
(MH "Validation Studies") OR (MH "Reliability and Validity") OR (MH "Instrument Validation") OR (MH "Internal Validity") OR (MH "Psychometrics") OR (MH "Measurement Issues and Assessments") OR TI ("psychometr*" OR "clinimetr*" OR "clinometr*" OR "Validation" OR (TI "construct" AND TI "valid*") OR (MH "Outcomes (Health Care)+") OR TI ("outcome assessment" OR outcome measure* OR "internal consistency") OR ((TI "cronbach*" AND TI (alpha OR alphas)) OR (TI "item" AND TI ("correlation*" OR "selection*" OR "reduction*")) OR TI ("agreement" OR "precision" OR "imprecision" OR "precise values") OR (TI intraclass AND TI correlation*) OR TI ("discriminative" OR "known group" OR "factor analysis" OR "factor analyses" OR "factor structure" OR "factor structures" OR "dimension*" OR "subscale*" OR (TI multitrait AND TI scaling AND (TI analysis OR TI analyses)) OR TI ("item discriminant" OR "interscale correlation*" OR "error" OR "errors" OR "individual variability" OR "interval variability" OR "rate variability") OR (TI variability AND (TI analysis OR TI values)) OR TI ("item response model" OR "irt" OR "rasch" OR "differential item functioning" OR "dif" OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence") OR (TI item* AND development*) OR AB ("psychometr*" OR "clinimetr*" OR "clinometr*" OR "Validation" OR (AB "construct" AND AB "valid*") OR AB ("outcome assessment" OR outcome measure* OR "internal consistency") OR ((AB "cronbach*" AND AB (alpha OR alphas)) OR (AB "item" AND AB ("correlation*" OR "selection*" OR "reduction*")) OR AB ("agreement" OR "precision" OR "imprecision" OR "precise values") OR (AB intraclass AND AB correlation*) OR AB ("discriminative" OR "known group" OR "factor analysis" OR "factor analyses" OR "factor structure" OR "factor structures" OR "dimension*" OR "subscale*" OR (AB multitrait AND AB scaling AND (AB analysis OR AB analyses)) OR AB ("item discriminant" OR "interscale correlation*" OR "error" OR "errors" OR

"individual variability" OR "interval variability" OR "rate variability") OR (AB variability AND (AB analysis OR AB values)) OR AB ("item response model" OR "irt" OR "rasch" OR "differential item functioning" OR "dif" OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence")

Appendix II. Characteristics of the included PROMs

PROM	Construct to be measured	Target population	Mode of administration	Recall period	(Sub)scale (s) (number of items)
Patient Satisfaction with Outpatient Physical Therapy Instrument[27]	Overall satisfaction with physical therapy	Adult outpatients with musculoskeletal conditions receiving physical therapy	Self-report	Undefined	Total number of items: 12 3 Domains: 1. Physiotherapist-patient interaction; 7 items 2. Clinic environment; 3 items 3. Convenience; 2 items
MedRisk Instrument to Measure Patient Satisfaction with Chiropractic Care (MRPS-CC)[28]	Overall satisfaction with chiropractic care	Adult outpatients with musculoskeletal conditions receiving chiropractic care	Self-report	Undefined	Total number of items: 13 3 Domains: 1. External factor; 6 items 2. Internal factors; 5 items 3. Global measures of patient satisfaction with care; 2 items
Physical Therapy Patient Satisfaction Questionnaire[29]	Overall satisfaction with physical therapy	Adult outpatients with musculoskeletal conditions receiving physical therapy	Self-report	Undefined	Total number of items: 20 11 Domains: 1. Treatment; 4 items 2. Privacy; 1 item 3. Convenience of appointment time; 2 items 4. Costs; 2 items 5. Billing; 1 item 6. Ease of scheduling an appointment; 1 item 7. Scheduling; 2 items 8. Wait time; 1 item 9. Courteous staff; 1 item 10. Physiotherapist courteous; 1 item 11. Overall satisfaction; 3 items

	Response options	Range of scores/ scoring	Type of measurement model	Original language	Available trans-lations
	5-point scale 1 (strongly disagree) to 5 (strongly agree)	Total range: 12-60 Physiotherapist-patient interaction: 7-35 Clinic environment: 3-15 Convenience: 2-10	Formative supplemented with reflective measures	English	
	6-point scale 1 (strongly disagree) to 5 (strongly agree) and a not applicable option	Total range: 13-65 External factors: 0- 30 Internal factors: 0-25 Global measures of patient satisfaction with care: 0-10	Formative supplemented with reflective measures	English	
	6-point scale 1 (strongly disagree) to 5 (strongly agree) and a 'no opinion' option	Total range: 0-100 Treatment: 0 -20 Privacy: 0-5 Convenience of appointment time: 0-10 Costs: 0-10 Billing: 0-5 Ease of scheduling an appointment: 0-5 Scheduling: 0-10 items Wait time: 0-5 Courteous staff: 0-5 Physiotherapist courteous: 0-5 Overall satisfaction: 0-15 items	Formative supplemented with reflective measures	English	Italian

PROM	Construct to be measured	Target population	Mode of administration	Recall period	(Sub)scale (s) (number of items)
Patient satisfaction with physical therapy instrument[30]	Overall satisfaction with physical therapy	Adult outpatients with musculoskeletal conditions receiving physical therapy	Self-report	Undefined	Total number of items: 23 5 Domains: 1. Patient-therapist interaction; 8 items 2. Physical environment; 4 items 3. Admission process, courtesy of the receptionist and support staff and waiting time; 6 items 4. Convenience; 2 items 5. Overall satisfaction; 3 items
Scale to measure patient satisfaction with physical therapy[31]	Overall satisfaction with physical therapy	Adult in- and outpatients with musculoskeletal conditions receiving physical therapy	Self-report	Undefined	Total number of items: 14 4 Domains: 1. Treatment; 5 items 2. Admission; 3 items 3. Logistics; 4 items 4. Global assessment; 2 items
Physical Therapy Outpatient Satisfaction Survey (PTOPS)[32]	Overall satisfaction with physical therapy	Adult outpatients with musculoskeletal conditions receiving physical therapy	Self-report	Undefined	Total number of items: 34 4 Domains: 1. Enhancers; 10 items 2. Detractors; 10 items 3. Location; 7 items 4. Cost; 7 items
Satisfaction with physiotherapy Questionnaire[33]	Overall satisfaction with physical therapy	Adult outpatients with musculoskeletal conditions receiving physical therapy	Self-report	Undefined	Total number of items: 16 5 Domains: 1. Contact; 3 items 2. Setting; 2 items 3. Intention; 2 items 4. Therapist; 7 items 5. Overall satisfaction; 2 items

* Positive scale (i.e., higher scores indicate greater satisfaction)

† Negative scale (i.e., higher scores indicate less satisfaction

	Response options	Range of scores/ scoring	Type of measurement model	Original language	Available trans-lations
	5-point scale 1 (terrible) to 5 (excellent) and 1 (never) to 5 (definitely yes)	Total range: 23-115 Patient-therapist interaction: 5-40 Physical environment: 4-20 Admission process, courtesy of the receptionist and support staff and waiting time: 6-30 Convenience: 2-10 Overall satisfaction: 3-15	Formative supplemented with reflective measures	Portu-guese	
	5-point scale 1 (poor) to 5 (excellent) and 1 (certainly not) to 5 (yes, certainly)	Total range: 14-70 Treatment: 5-25 Admission: 3-15 Logistics: 4-20 Global assessment: 2-10	Formative	French	English
	5-point scale: 1 (strongly disagree) to 5 (strongly agree)	Total range: 34-170 Enhancers*: 10-50 Location*: 7-35 Detractors†: 10-50 Cost†: 7-35	Formative supplemented with reflective measures	English	
	No information provided	No information provided	Formative supplemented with reflective measures	English and Arabic	

PART II

Advanced practice physiotherapy within Dutch primary care


4

The introduction of advanced practice physiotherapy within Dutch primary care is a quest for possibilities, added value, and mutual trust: a qualitative study amongst advanced practice physiotherapists and general practitioners

BMC health services research. 2022; doi.org/10.1186/s12913-022-07906-6

Pellekooren S, Donker M, Reijnders E, Boutalab L, Ostelo RWJG, van Tulder MW, Pool-Goudzwaard AL.

Abstract

Background

An ageing population and an increasing number of chronically ill patients does not only lead to more healthcare utilization within general practices, but also to more complex healthcare demands. To cope with this, GPs organise support and expertise within their practices by collaborating with other healthcare providers. One of these healthcare providers is the Advanced Practitioner Physiotherapist (APP), also referred to as Extended Scope Practitioners (ESP), who take over tasks in musculoskeletal care traditionally performed by GPs. Despite the increased deployment and added value of Advanced Practitioner Physiotherapy (APP) in internationally, APP is not yet widely accepted within Dutch primary care. This may be due to specific constraints in the implementation of APP within the Dutch healthcare system.

Objective

To explore the experiences and perceptions of Advanced Practitioner Physiotherapists (APPs) and General Practitioners (GPs) with respect to implementing APP within Dutch primary care.

Methods

This explorative and interpretive qualitative study included 12 APPs and 3 GPs who were in various stages of implementing an APP care model. Semi-structured interviews were conducted between January and March 2021. The topic list was based on existing literature, the personal input of researchers, and the Constellation Approach framework. Data were analysed using a thematic inductive approach.

Results

Four main themes emerged from the data; 1) Both GPs' trust in APP and a clear added value of APP are critical for starting implementation, 2) APPs need continuous support from GPs, 3) APPs believe that their position needs strengthening, and 4) Implementation of the APP model creates tension over ownership. These four themes highlight the perceived difficulties in gaining trust, lack of clarity over the added value of APP, ambiguity over APPs' professional profile and positioning, a need on behalf of GPs to maintain authority, lack of reimbursement structure, and the struggle APPs face to strike a balance with current care.

Conclusions

This study demonstrates that implementing an APP model of care is challenging, in part, because the deployment of APP does not sufficiently align with the core values of GPs, while GPs appear reluctant to hand over control of elements of patient care to APPs. APPs do not appear to have ownership over the implementation, given their strong dependence on the practice, values and needs of GPs.

Introduction

Healthcare utilisation has steadily risen over the years and is expected to increase even further as a result of healthcare innovations and an aging population.[1] Similar to other healthcare systems in the world, the Dutch healthcare system faces the challenge of how to deal with the increased demand for care, which, in turn, increases the workload for healthcare workers. Dutch general practitioners (GPs) in particular, who are taking over tasks from secondary healthcare facilities while, simultaneously, seeing a decline in the number of their colleagues,[2] have reported a considerable increase in their workload.[3]

One way to reduce the workload of GPs is to deploy additional care givers, such as, for example, nurse practitioners and physician assistants, who have taken over the delivery of care for patients with chronic conditions like diabetes[4] and depression.[5-7] In light of these examples, there have been various initiatives across the globe to relieve GPs of having to care for the significant group of patients with musculoskeletal conditions seeking treatment, via the use of Advanced Practitioner Physiotherapists (APPs), who are also referred to as Extended Scope Practitioners (ESPs).[8,9]

APPs operate at an advanced level of practice and provide care that is traditionally provided by other medical professionals, and are responsible for setting and communicating diagnoses, triaging for surgery or surgical opinions, ordering diagnostic imaging or laboratory tests, and prescribing/injecting medications.[9,10] A recently published study showed that the deployment of APPs contributes to the accessibility of care with comparable health effects, diagnostic accuracy, and patient satisfaction.[9] In addition, the Advanced Practice Physiotherapy (APP) model of care has been shown to result in lower healthcare costs compared to usual care.[11] Based on these international findings, APP thus appears to be an appropriate alternative in treating patients with musculoskeletal conditions, which, in turn, may help reduce GPs' workload.

In response to these international developments, APP was introduced in the Netherlands a few years ago via ESPs. This name was in accordance with extant literature at that time, which described the substitution of medical care by a physiotherapist as an ESP. However, the word practitioner was replaced by specialist to emphasise the difference between a physiotherapist qua practitioner, who treats patients via regular care, and a specialist who has more extensive tasks and responsibilities.[8]

Despite promising results internationally, the deployment of APP within Dutch primary care has yet to garner wide acceptance. This may derive from barriers in the implementation of such innovations within the Dutch healthcare system, which is

characterised by a demand-driven system with regulated competition and elements of both public and private insurance. All residents are entitled to a comprehensive health insurance package. This so-called basic health insurance package is compulsory, and the reimbursement structure is determined by the government. In addition to the compulsory insurance package, residents can also opt for supplementary insurance for forms of care that are not reimbursed by basic health insurance, such as, for example, physiotherapy. Within primary care in this healthcare system, there is a central role for the strongly positioned GP,[12,13] who not only functions as the gatekeeper for secondary care but also serves as the fixed first point of contact as all residents are registered with their own GP. Consequently, there is a long-term relationship between patients and GPs, allied with a strong focus on shared decision-making and high continuity of care.[13]

Several qualitative studies[14–18] have explored the barriers and facilitators to the implementation of APP within secondary healthcare settings, concluding that knowledge, skills, availability of APP, motivation, and experience all have a large impact upon its successful implementation. Although these studies are undoubtedly of great value in terms of improving the implementation of APP, their outcomes are not necessarily applicable to the Dutch primary care setting. This is because these studies focused on the implementation of APP within secondary care facilities in other healthcare systems, and, as such, those barriers and facilitators that are specific to the Dutch primary care setting may not have been addressed. Therefore, this study aimed to explore the experiences and perceptions of Dutch APPs and GPs in implementing APP within Dutch primary care. Although the term ESP is used in the Netherlands, the acronym APP(s) is used throughout the manuscript, hereby following the term used in recent publications.[8,9,11]

Method

Design

This was an explorative and interpretive qualitative study among Dutch APPs and GPs who were in various stages of implementing a musculoskeletal APP care model within a primary care setting. Semi-structured interviews were carried out between January and March 2021.

Framework

The framework used to enrich the interview guide was the Constellation Approach,[19] which was developed to analyse transitions in complex systems. The constellation approach assumes that complex societal systems, such as, for example, the Dutch

healthcare system, consist of several subsystems, or so-called constellations. These constellations interact with, and adapt to, each other as well as their system's environment. Each constellation comprises three elements, namely structure, culture, and practice. Structure refers to the physical, economic, financial, legal, organisational, and power structures that influence the behaviour of actors within a constellation. Practice concerns the actual actions that are undertaken within a constellation, such as the interactions between healthcare professionals and patients or between professionals and management. Culture pertains to the set of values, perceptions, and interpretations of actors within the constellation (e.g., patients, physicians, healthcare providers and insurance companies). While a constellation may change as a result of a variety of influences originating from within the organisation, it can also be demand- or supply-driven. When implementing APP in the Netherlands, the critical question is whether the healthcare structure, the beliefs of those parties involved, and the daily practice of musculoskeletal care are open to the introduction of APP.

Participants

In order to achieve the maximum degree of variation and collect meaningful experiences and perceptions, participants were recruited via different channels. APPs were recruited through the researchers' existing network, which consisted of a group of approximately 30 APPs who were already known to SP as part of an observational pilot study. In parallel with this, participants were also recruited through both alumni and professional associations to include participants that were in other phases of implementation. These APPs were contacted via social media and through a call in the newsletters of the professional association. APPs were selected based on the stage of their implementation (e.g., start-up phase or established practice), personal characteristics (e.g., years of experience, region and attended training), and practice characteristics (e.g., self-employed or embedded in GP practice). GPs were recruited through the researchers' existing network and regional GP associations and were approached by phone or email. On the whole, the willingness of GPs to participate was limited since these GPs were unfamiliar with APP, did not endorse it as a model of care, or indicated that they had no difficulties in providing care for patients with musculoskeletal complaints and were, therefore, not interested in the topic. As GPs proved to be a difficult group to recruit, they were selected based on convenience sampling. Twelve APPs and four GPs were included in the study, of which one GP subsequently decided not to participate due to their busy schedule.

Data collection and Data analysis

The topic list was based on the available knowledge from extant literature,[14–18] before then being enriched with elements of the constellation approach (i.e., culture, structure, and practice characteristics)[19] and the personal input of the researchers (SP, ER, AP, RO, and MvT). SP's personal input resulted in the inclusion of topics related to the APP perspective, ER's input led to the inclusion of topics focused on GPs' perspective. Subsequently AP, RO and MvT checked the topic list for completeness and the neutrality of the questions. The final topic list included the following: reasons for starting a collaboration, extent of implementation and activities of APPs, training of APPs, awareness and need for APPs, support from the professional association, barriers and facilitators in the collaboration, alignment with the core values of Dutch general practice, and opportunities and future prospects for APPs.

The interviews lasted around 60 minutes, with the exception of one interview that lasted thirty minutes and were conducted via an online video call. The interviews were audio recorded and fieldnotes were taken. A pseudonymised verbatim transcription of the audio recordings was obtained. Summaries (i.e., member checks) were sent to the participants to provide them with the opportunity to comment and adjust the summary of their interview.[20] As part of this process, we stressed that the summary was the researcher's interpretation of the interview and that any changes or additions were welcome. Five of the participants had some small remarks for clarification, which were accounted for during the further analysis. After reading their summary, some APPs expressed their disappointment and frustration toward the lengthy and cumbersome process when they became aware of the actual level of implementation. This, in turn, influenced the analysis and the subsequent development of themes.

Data were thematically analysed by means of an inductive approach.[21] Two researchers (SP and LB) familiarised themselves with the data by reading and rereading the transcripts, before subsequently independently selecting relevant fragments from three interviews by assigning open codes. These codes were then discussed and agreed upon with a third researcher (MD). After coding the first three interviews, a set of open codes were composed, which the remaining interviews were then coded with. Within this set, it was possible to add new codes. Any new codes were discussed within the research team and the set of codes were then adjusted accordingly, if necessary. Next, the codes were compared and grouped according to main- and subthemes. After interviewing nine APPs, no new themes were found from the APPs' perspective. However, we did decide to conduct an additional three interviews with APPs to explore the GP perspective further, as we felt that the GP perspective was less reflected as a result of having only conducted three interviews with GPs. We found that in prior

interviews with APPs, APPs had also put forward elements of the GP perspective, so we thought that additional interviews could help contribute to a better understanding of the GP perspective. However, no new themes emerged from these additional interviews. Valuable quotes were selected during the analysis and then discussed and interpreted among the researchers (SP, MD). All analyses were carried out in MAXQDA (version 2020).

Ethical considerations

All participants provided their informed consent prior to participating in the study. Ethical approval was obtained from the Medical Ethics Committee of VU University Medical Centre in Amsterdam; reference number 2020.17.

Results

The APPs who participated in the study differed in terms of their personal characteristics and specific working conditions, such as, for example, their forms of reimbursement and level of organisational embedment. Seven of the twelve APPs collaborated with a GP, of which two worked under the supervision of the GP and five worked independently. More details on the participants can be found in Table 1.

Four main themes derived from the data; 1) Both GPs' trust in APP and the clear added value of APP are critical for starting implementation, 2) APPs need continuous support from GPs, 3) APPs believe that their position needs strengthening, and 4) Implementation of the APP model creates tension over ownership. Details on the derived subthemes and axial codes can be found in the code tree, which is presented in Table 2.

1.Both GPs' trust in APP and a clear added value of APP are critical for starting implementation

GPs need to trust APP

All the APPs indicated that having a long-term relationship with a GP is a prerequisite for introducing an APP model of care. GPs need to trust in both the competencies and motivations of APPs in order to develop confidence in the collaboration and eventually hand over care delivery to APPs. This trust can be built by working together. Those APPs that lacked such a pre-existing relationship experienced difficulties in connecting with GPs, gaining their trust, and introducing an APP model of care without calling into question GPs' competences.

The most important factor is trust. Trust that those who are doing the project, APP X and APP Y, are competent in the matter. That they are also prepared to behave in this way, and not say, this is a disguised way of bringing in more clients at the end of the day, so that is the most important thing, I think. [GP 1]

GPs doubt added value of APP

Many GPs are still unfamiliar with APP and what the profession precisely entails. GPs find it difficult to see both the added value of APP over specialised physiotherapy and how an APP model of care would improve their current practice.

Because of course there are so many different therapists with all kinds of functions. It has to be very clear what exactly the added value is for us to refer a patient to an APP instead of a 'regular' physio. [GP 2]

In addition, the GPs indicated that it remains unclear what an APP model of care offers them personally and professionally, whether it be in terms of time savings or better quality of care. All GPs reported that one-off assessments by APPs would undoubtedly contribute to greater musculoskeletal expertise within their GP practice, while one GP mentioned that potentially reducing their own workload gave it added value. However, two of the three GPs interviewed indicated that heavy workloads were not primarily caused by patients with musculoskeletal problems, but by the relocation of care from secondary care settings, such as, for example, mental healthcare facilities and care for the elderly.

And the problems with the elderly are just very heavy, when you have so many elderly. I have a lot of elderly people, and they all live at home, and I have a lot of demented people, and there is little home care. It is a familiar story. Not enough places, they cannot be admitted, or do not want to be admitted. That is what takes up most of my time. That will continue to be my practice. So that is where I need the most support actually. [GP 3]

Table 1. Characteristics of the participants

Respondent	Gender	Age	Years of work experience as PT	Years after graduating APP	Currently practicing APP	APPs working under supervision
APP 1*	Female	> 50	> 25	> 2	Yes	+**
APP 2*	Female	> 50	> 30	> 2	No	+**
APP 3*	Female	> 60	> 30	> 2	Yes	+
APP 4*	Male	> 50	> 20	> 2	Yes	+
APP 5	Female	> 35	> 10	< 2	Yes	- ***
APP 6*	Female	> 45	> 20	< 2	Yes	- ***
APP 7*	Male	> 55	> 35	> 6	Yes	+
APP 8*	Female	> 40	> 20	> 2	Yes	+
APP 9*	Female	> 40	> 15	> 4	No	Na
APP 10*	Male	> 30	> 10	> 2	No	Na
APP 11*	Male	> 40	> 15	< 2	No	Na
APP 12*	Male	> 40	> 20	< 2	No	Na

Respondent	Gender	Age	Years of work experience	Practice composition	Number of patients registered to GP practice	Number of collaborating APPs	APPs working under supervision
GP1*	Male	>55	25	1 GP, 1 permanent alternate	2200	2	+
GP2*	Female	>40	13	2 GP	2900	1	-
GP3*	Female	>50	21	1 GP, 1 HIDHA, 1 HAIOS	3000	1	+

APP = Advanced Practitioner Physiotherapy; APPs = Advanced Practitioner Physiotherapists; GP = General Practitioner; PT = Physiotherapist; HIDHA= GP employed by another GP; HAIOS = GP in training; * = owner practice; ** joint consultation; *** independent consultation.

Table 2. Code tree

Themes	Subthemes	Axial codes
Both GPs' trust in APP and a clear added value of APP are critical for starting implementation	GPs need to trust APP	
	GPs doubt added value of APP	
APPs need continuous support from GPs	APPs need the full commitment of GPs to start	APPs cannot refer to secondary care on their own
		Limited availability of patient information
		Triaging patients lacks criteria
		APPs and GPs want to scale-up
		GPs and APPs struggle with who is in charge of the care pathway
	APPs require support from GPs while they build-up their self-confidence	Insecurity during delivery of care
		Insecurity during team interactions
		More work experience increases their self confidence
	Establishment of proper reimbursement is crucial	
APPs believe that their position needs strengthening	GPs want to retain their authority and control	Competencies and attainment levels are poorly crystalised
		Different preferences for type of employment and final responsibility
		APPs experienced tension between GPs' standards and their working methods
	More guidance from the professional association is desirable	APPs want more backing from trade organisation
		Trade organisation needs to be a driving force towards stakeholders
	APPs found limited added value in the training they attended	Work experience influences the added value of the training
		Curriculum needs more in-depth and practical training

Table 2. Continued

Themes	Subthemes	Axial codes
Implementation of the APP model creates tension over ownership	No place for APP among physiotherapy yet	Gaining trust amongst physiotherapists with whom they need to collaborate
		Controversy over the positioning of APPs
	Finding the balance between taking over GP care and safeguarding core values	Deployment of APP jeopardises patient-centred care
		Ensuring the independent delivery of care appears to be an unfeasible ideal
		GPs must be able to maintain the delivery of general medical care at a qualified level
		APPs and GPs need to develop a common language

2. APPs need continuous support from GPs

APPs need the full commitment of GPs to start

The vast majority of APPs indicated that the start-up stage of an APP model of care is a long process that involves many steps, especially for APPs that are not embedded in the GP practice, who also must deal with legislative issues like the General Data Protection Regulation and doctor-patient confidentiality. The APPs indicated that receiving support from GPs is essential for referring to secondary care, eliciting enough patient information for setting out the care pathway properly, and for setting up referral streams. However, some APPs experienced that GPs tend to be less committed in implementing an APP model of care since the interest mainly lies with APPs. Although embedding APP within GPs' practice can overcome some of these aforementioned hurdles, it is not attractive to all GPs because it means taking on more staff.

And I can only speak for my own GPs, something I've discussed a lot over the last year, GPs don't want to grow in the size of their practices either, they're not waiting for 30 practice support staff. The role that we have now is actually quite fine, nice, I don't have anything to do with you, I don't have to take care of you when you're sick, you take care of it there, we take care of it here, that's what these GPs like very much. And my GPs are not waiting for APP to come in as well. [APP 4]

APPs require support from GPs while they build-up their self-confidence

All practicing APPs sometimes feel insecure and vulnerable over having primary responsibility for patients' wellbeing, especially when their complaints may not appear to be related to the musculoskeletal domain. Having consultations with fellow APPs or the authorising GP helps to reduce this uncertainty. Practicing APPs expressed feeling uncertainty when reporting to GPs and felt that they were not allowed to make mistakes in the initial stage where they still had to prove themselves. Their self-confidence would grow by receiving positive feedback from GPs and gaining more work experience.

It would be a death blow of course, everyone makes mistakes, but it would mean the end of everything if we had a lot of misdiagnoses in the initial phase. Then, immediately, seeds of doubt are sown, and of course, we cannot have that. [APP 4]

APPs needs practical support from multiple GPs to carry out their practice

Some APPs and GPs indicated that a uniform way of working, and communication are paramount for both ensuring high-quality care and for carrying out joint consultations to this end. All APPs and GPs preferred a workplace within a health centre where several GPs work, because APPs are then embedded in the GP practice and short lines of communication are established. However, this is difficult to realise in practice due to the lack of working space within most health centres. A few APPs stated that working out of one's own physiotherapy practice is attractive, as this increases the referral of patients for physiotherapy treatment, and, as such, one's income. The GPs indicated that working out of one's own physiotherapy practice is not desirable, as the independence of care and the role of APPs then comes into question. All participants saw the added value of scaling up the team, as far as this ensures continuity, independence, and quality of care. All GPs indicated that they struggle with referring a sufficient number of patients and are uncertain over which APP they should be contracting. A few APPs mentioned that it is difficult to scale up due to both the insufficient number of trained APPs in their work area and the competitive attitude of other APPs.

I think that in our case she [APP] should actually work for several practices, because one practice – even though I have a large practice – one should have more opportunities available. You always have people who think, I would rather go to the GP because then I will see the doctor again, too. Or imagine, you have already been through a lot with a patient and then the patient prefers the GP. Not that it is necessarily better in terms of content, but because the GP is a trusted figure. [GP 3]

Establishment of proper reimbursement is crucial

All APPs and GPs indicated that the lack of an appropriate financing structure is a major barrier for APPs, GPs, and patients. Although reimbursement is possible through the health insurer's innovation fund, GPs are either not able or are unwilling to utilise this. As patients are used to GP care being reimbursed from their public health insurance, GPs would not only have to convince patients of the added value of an APP over a physiotherapist, but also inform them about the additional costs. These costs would then either be paid out of patients' own pockets or at the expense of the number of physiotherapy treatments covered by their supplementary insurance. Some GPs indicated that they perceive this restricted accessibility of care based on a patient's financial position as unpleasant and/or unethical.

GPs were not really keen on using funds from the innovation fund of the health insurers for this purpose. Many GPs had also just made additional investments in physician's assistants. So that was an issue. Also, because we have another group of GPs here, some of whom think that extended scope is unnecessary. [APP 10]

3. APPs believe that their position needs strengthening

GPs want to retain their authority and control

The APP competency profile developed by the Dutch professional association for APP is unknown to many APPs and leaves room for differentiation in the function of APP. All of the participating APPs had different views on competencies, end terms, tasks, patient population, and their position in the care pathway. Some APPs indicated that this flexibility in their profile leads to ambiguity and confusion amongst GPs and patients. There is no consensus yet amongst both APPs and GPs over the establishment of employment of APPs within the GP practice and if APPs should work according to GP professional standards. Moreover, the professional role does not only dependent on the professional profile. Some APPs who do set out the care pathway themselves indicate that, despite agreements made, they sometimes have trouble staying in charge of the treatment plan, as in practice their role is also influenced by old behavioural patterns of patients and GPs. In addition, most APPs argued that their role as APP seems not only to be determined by the professional profile but also by the extent to which APP is allowed to work next to the GP by the GP. The APPs also indicated that they are cautious in taking over too much care at the one time and proceed step by step to avoid resistance from the GP.

Initially that would not matter to me. I think that we should say that, as a goal, it will eventually be fully under APP own authority. Certainly, to get the GPs on board I think that you must first do this under the GP's authority, until they themselves

conclude, no, you can do this on your own just fine, and I don't need to be behind this, like some version of extension of care. So, I think that this must be introduced step by step. In particular if you also notice that they [GPs] are going to get up in arms, then you should introduce that very slowly. And prove yourself first. You must. [APP 6]

Although the APPs indicated that they are willing to temporally work under GP authority, two GPs stressed that they have no intention of handing over full authority. Rather, they stated that they will either opt for joint consultation or deploy APPs under supervision and set out the care pathway themselves, thereby retaining control.

I should like it to be under my supervision because I think that in this way I can offer an extra service to my patients, a broader selection of diagnostic skills and I do not throw this [treatment responsibility] out. So, for as far as this goes, I want them [patients] to go to it [APP], and then they often return to me, and we discuss what the proposed treatment plan is. In this way I do not let go of them. [GP1]

More guidance from the professional association is desirable

Almost all APPs stated that they missed the support of a professional association when starting their APP practice. That is to say, they missed having a platform to fall back on and get more guidance, such as, for example, a concrete plan of action, standard documentation, and advice on how to communicate with GPs, which was needed but not yet available. Virtually all the APPs felt that the professional association is not sufficiently visible to the various stakeholders, while developments within the professional association take a long time. All APPs indicated that the implementation of APP would benefit from a decisive board that is actively engaged in creating support amongst stakeholders. The lack of direction from the professional association leads to many individual initiatives, loss of control over this growing profession, and differences in the interpretation of the role and working method of APPs.

I understand that as well, because it is a new association and must be built from the ground up. Furthermore, it is not their main task, they also have of course their own jobs to do. But certainly, for this project, things [documentation] have been agreed upon and were to have been sent in, but this has not happened, which is a pity, because as a pioneer, you really need support. And that is not happening. Or at any rate, too little. [APP 2]

APPs found limited added value in the training they attended

The vast majority of the APPs interviewed said that the training they had undergone contributed little to the knowledge and skills they had already acquired in either their work as manual or sport physiotherapists or in their previous master's degree courses. Some APPs indicated that, compared to other countries, the scope of the training was too limited, and that practical education under the supervision of a doctor was lacking.

This is fine for a few weeks, going a bit deeper into things, but does not compare with the role they play abroad, nor the training they receive for this.... They have had a completely different training in this, and this I think, is what is keeping us from getting any further with this APP story in the Netherlands. [APP 9].

4. Implementation of the APP model creates tension over ownership

No place for APP among physiotherapy yet

All the APPs indicated that building a collaborative network with physiotherapists in their region costs them lots of time and effort, as the concept of APP is still relatively unknown. Feelings of anxiety over losing patients as well as unfair competition amongst physiotherapists both contribute to the slow acceptance of APP, despite the efforts of APPs themselves to stress that it is not their intention to treat patients themselves. Some APPs reported that with the current reimbursement APP acts as a competitor to physiotherapists, which has a deleterious impact upon their cooperation.

How do I notice this happening? Not providing information, not sharing patients, getting angry with you the moment you see a patient and call about it, or do a report, or have an other idea. If you want to set up a project about APP care, and you go to a big player in the neighbourhood who also has a similar plan, something broader, and you say, well, let us join forces, then it is all impossible. No, it is all too sensitive, too much me, me, me.... This leads to extremely unpleasant conversations. [APP 1]

Some APPs mentioned that combining the APP care model with direct access physiotherapy results in APP functioning as an additional gatekeeper along with the GP. This may be used as a unique selling point to expand one's own physiotherapy practice and make more money, which, in turn, leads to feelings of unfair competition and resistance towards APP. One APP stated that there are ongoing discussions both in the field of work and at the management level who can be an APP and who cannot. Some APPs said that they had experienced that some physiotherapists present themselves as APPs without undergoing the proper training. Indeed, one APP even mentioned that the Royal Dutch Society for Physical Therapy (KNGF) agrees that at least in principle, every physiotherapist can carry out APP.

The other one, practice X, just wants to scale-up. And they also want to be a part of it [setting up an APP practice in the region], but then it is no longer about the content. The worst thing I found, was that nobody has done training in APP, but they pretend to be on top of it... I think the Society, that is the regional representative of KNGF, believes that every physiotherapist should be able to be an APP. I do not agree with him at all. Manual therapy and sport physiotherapy may think so, but the KNGF has a completely different opinion. At least in our region, the KNGF simply airs this. This is already a difficult matter. [APP 3]

Finding the balance between taking over GP care and safeguarding core values

All the GPs indicated that collaborating with APPs may jeopardise patients' interests, due to a restricted choice of care provider and further fragmentation of care. All APPs and GPs endorsed that APP should operate as an independent point of care and emphasised that one should not position APP as part of the business model of one's own physiotherapy practice. However, the APPs indicated that this independence is difficult to realise as both APP care and physiotherapy care are typically provided alongside each other, due to the limited number of patients, lack of workplace at healthcare centres, and poor understanding of APP services by patients. Most APPs stressed that providing independent care still has a long way to go and may in fact not be feasible, especially for those APPs that are affiliated with large physiotherapy practices that provide a wide variety of in-house treatment options.

On the other hand, I discussed this [lack of independent care delivery], with fellow physios already during my training, and they all say, are you crazy, everyone works that way within primary care. And they all pass the buck to each other. So, I let it rest for a while. They are right, I think the same way, but that is partly a hypocritical remark for everyone. So, then everyone needs to put his own house in order, and then we can all be morally justified. But to be honest, because I am quite a moralist, if I let go of that, I think it is going to be a difficult issue. I agree, I totally agree, I think that is the way it should be, in the ideal world, but I think we are a long way from that. [APP 11]

All the GPs mentioned that in order to provide proper general medical care, gaining and maintaining experience with musculoskeletal complaints is absolutely essential. Some APPs and GPs indicated that not all GPs are willing to hand over patients with musculoskeletal complaints due to their personal interest in this population and/or beliefs about the content of their profession and Dutch GP core values. Indeed, two out of three of the GPs interviewed felt that APPs still have to grow into the culture of the GP practice and find a way to connect with the core values. Some APPs noted that connecting with the GPs and relating to the mutual dynamics of GPs can be difficult due to other perspectives on the quality of care.

Totally different, if you think it might be a good idea to involve a secondary care orthopaedic, then the GP says, oh, no, you mustn't, because that is seen as primary care in disguise. So, you definitely should not do that! You are just not aware of all these strategically sensitive things. And you think you have a great product, and the GP thinks, how so? I do not need you at all. So how are you going to connect with them? [APP 1]

Discussion

This study explored the experiences and perceptions of APPs and GPs towards both the implementation and deployment of APP within Dutch primary care and found that it is difficult for APPs to carve out a place for themselves within the healthcare landscape.

Within the present study, four themes emerged from the data through which APPs and GPs' experiences of APP deployment and implementation can be understood. The first theme sheds lights on the fact that the success of APP depends on both the trust of the GP and whether they perceive it as having clear added value in comparison to the usual care. The second theme underscores that the support of GPs is essential for APPs, as far as it helps to, amongst other things, get different referral flows going. The GP also plays an important role in terms of building the self-confidence of APPs, in creating uniformity within patient care, and in terms of helping to bring about a team that works under one roof. The lack of funding for APP raises concerns over the deployment of APP among APPs, GPs, and physiotherapists. The third theme points towards the fact that the position of APPs needs strengthening. Indeed, the professional profile of APP is something that proved to be unclear to both GPs and APPs themselves. In the absence of a uniformed way of working, everyone is still searching, which, in turn, results in diversification. GPs' reluctance to hand over control also profoundly impacts on the role of APPs. Amongst APPs, there is a need for better positioning, support, and profiling from the professional association as well as for training which includes more depth and practical education. The fourth theme pertains to both the tension that persists around ownership of patients with musculoskeletal complaints and the competition between APPs and physiotherapists. This is compounded by a lack of adequate funding and the ability to generate patient flow for the physical therapy practice to which APPs are affiliated. Moreover, the APP model seems to insufficiently adhere to GPs' core values.

Comparison with literature

Many of the themes identified are in accordance with earlier publications on APP, such as the role of trust and need for acceptance by doctors,[14–16] recognition of the added value by doctors[14,16,17] and the establishment of an appropriate financing structure.[14,17] The present study shows that many of these previously identified factors, such as physician trust and demonstrating clear added value to stakeholders and the financing structure, have hitherto not been sufficiently realised to facilitate the implementation of APP within Dutch healthcare. The most important barrier, however, appears to be GPs' reluctance to hand over authority and control. This appears to stem from specific characteristics of Dutch general practice, such as long-term doctor-patient relationships and GPs' strongly held core values, but also derives from the traditional authority that GPs have over physiotherapists as a result of differences in educational level, which persists because of the lack of sufficient training and entrusted professional activities.

Introduction and support from within the organization have been described as helpful in studies of APP embedded in secondary or tertiary care.[14,15,17,18] Such support is lacking in the implementation of APP in the Netherlands, and individual APPs working independently in primary care must build a partnership without any support.

A number of studies have shown that the availability of training at an appropriate level is critically important.[14–16] Our study shows that, according to the experiences of the APPs, both the form and scope of the current education is not in line with the demands of the professional field and, moreover, is not sufficiently different from their prior training and thus lacks added value for them. In addition, individual APPs are currently responsible for organising their own practical training in the field. It is unclear to what extent this is feasible for APPs given the limited scale of most of their collaborations, where guidance often has to be provided by an individual GP, while gaining practical experience is dependent on the limited number of patients registered with this GP.

Furthermore, a number of studies have shown that a clear delineation of the role of APPs and greater standardisation of working procedures is important.[14,17,18] A recent qualitative study examining the goals, roles and tasks of APPs in the Netherlands revealed that the participants found it difficult to state clear goals for APPs and that there is no consensus concerning the positioning of APPs.[22] A study on how best to shape the interprofessional collaboration between GPs and established healthcare professionals[23] showed that these collaborations do not always go well and that it is crucial to establish a shared vision and clarity over work structure, procedure, and role distribution. Awareness of each other's context and expectations was also found to play

a key role. According to the APPs and GPs who took part in this study, a clearly defined role and standardisation of process and working methods of APPs has yet to be realised. This makes it incredibly difficult to develop the partnership between APPs and GPs.

Amongst GPs, there is a need to improve the already existing collaboration with physiotherapists to ensure the increasingly complex care of patients with musculoskeletal complaints.[23] Within current Dutch primary care, around half of all GPs already have an existing collaboration with a physiotherapist,[24] while a large proportion of patients with musculoskeletal complaints visit a physiotherapist via Direct Access Physiotherapy.[25] In this context, the question is whether there is a need therefore for a new type of care provider, such as APPs, or whether there is a need to revise the existing collaborations with physiotherapists, by improving the level of communication and having one-off diagnostic consultations.

In other countries, such as Australia and the United Kingdom, APP has emerged in response to urgent demand from physicians.[14,16] Here, involved stakeholders have felt sufficient urgency to change and, moreover, physicians have endorsed the need for the use of APPs.[14–16] Within the present study, there was no such urgency and need expressed by GPs. This might relate to differences in the organisation of healthcare systems, not to mention the good accessibility and continuity of Dutch GP care. It has also been found that when APP is not initiated by physicians themselves, then its implementation is altogether more difficult and dependent on goodwill.[16] This also appears to be the case with the implementation of APP in the Netherlands.

It remains to be seen to what extent APP fits within the Dutch College of General Practitioners future vision[26] in which the GP, as the first point of contact, maintains an overview of medical care and determines, together with the patient, what care is necessary and appropriate. The Dutch General Practitioners Association has recommended that, when entering a partnership with a new care provider, GPs must determine, before doing so, to what extent the core values and core tasks are to be guaranteed.[27–29] Moreover, GPs are advised to assess if the collaboration with this new care provider corresponds to their own preferences, ambitions, and vision of GP care. [27–29] In addition, a study amongst patients of Dutch GPs showed that patients' wishes regarding healthcare providers should be considered in ever-increasing collaborations with the GP practice.[30] At present, it is not feasible for APPs to adequately align with the key conditions that GPs want to see fulfilled before they are willing to change their practices, while it remains unclear to what extent patients' wishes are being heeded in the implementation and deployment of APP.

The importance of connecting to core values was also highlighted in a study evaluating barriers to the implementation of the Dutch General Practitioners Association treatment standards.[31] This study demonstrated that, despite the positive attitude of GPs towards the implementation of these standards, GPs only follow the standards when they are in line with the core value of patient-centred care. This makes it clear that, even with an improved positioning of APP, connecting to the core value of person-centred care is decisive in successfully implementing APP. There seems to be a lack of vision regarding under what conditions this can be met, which, in turn, makes it difficult for individual APPs to connect with GPs.

Strengths and Limitations

One of the strengths of this study is its credibility.[32] The starting point was an extensive literature review, which subsequently formed the basis of the interview guide. Multiple researchers collaborated on this study, and during the analysis, two researchers coded independently of each other, and subsequently the codes and themes were extensively coordinated and discussed within the research team. In addition, the full scope of the use and implementation of APP was examined by using concepts from the constellation approach as sensitising concepts in developing the interview guide. Moreover, all the participants were sent a member check after the interview and their responses were included in the analysis. Another strength concerns the conformability[32] of the results, as a large team from different backgrounds worked on the study. Moreover, a good audit trial was carried out, during which the selection process around the analysis was recorded and explicit attention was paid to the views and thought processes of each individual team member. This was an important aspect as one individual researcher (SP) is a physiotherapist and was involved in conducting an observational pilot study that evaluated the APP model of care and, as such, was more familiar with the perspective of APPs. The presence of possible disconfirmatory cases was discussed within the research team, but although there was diversification amongst the participants, no disconfirmatory cases were identified. The findings were in line with other studies examining the implementation of APP models of care. The transferability[32] of the findings is unclear. Despite there being similar findings in extant literature on implementation level, comparison with international literature is difficult given the specific Dutch context. Although we used maximum variation sampling, we were compelled to recruit GPs through convenience sampling given the limited number of GPs who were willing to participate, which meant that we failed to include GPs who were not open to implementing the APP model. This probably hinders the transferability of our findings, as far as we may have missed aspects of the GP perspective. However, gaining trust in APP, the need for a clear added value, reluctance to hand over control, and strongly held core values was expressed by all the participating GPs. There may

also be shortcomings in the dependability[32] of the findings. Although we collected data until no new themes derived and flexible analysis took place, data collection and data analysis were not a wholly iterative process. In addition, there is a possibility that some of the participants may have felt less free to express themselves during the interview, out of concern that they may have, despite being anonymised, been recognised by colleagues and stakeholders based on their specific characteristics.

Conclusion

The results of this study show that implementing an APP model of care is challenging within the Dutch healthcare system. The deployment of APP does not sufficiently align with the core values of GPs, and GPs appear to be reluctant to hand over some control over patient care to APPs. Therefore, APPs do not appear to have ownership over the implementation, given their strong dependence on the practice, values and needs of GPs.

References

1. Garson Jr A, Levin S. Ten 10-year trends for the future of healthcare: implications for academic health centers. *Ochsner J*. 2001;3:10-15.
2. van der Horst H, Bindels P, Assendelft P, et al. Hoogste tijd voor minder patiënten per huisarts. *Huisarts Wet*. 2018;61(3):30-31. doi:10.1007/s12445-018-0056-9
3. Schäfer W, van den Berg M, Groenewegen P. De werkbelasting van huisartsen in internationaal perspectief. *Huisarts Wet* . 2016;59(94):94-101.
4. Laurant M, van der Biezen M, Wijers N, Watananirun K, Kontopantelis E, van Vught AJ. Nurses as substitutes for doctors in primary care. *Cochrane Database of Systematic Reviews*. 2018;2019(2). doi:10.1002/14651858.CD001271.pub3
5. Bennett M, Walters K, Drennan V, Buszewicz M. Structured Pro-Active Care for Chronic Depression by Practice Nurses in Primary Care: A Qualitative Evaluation. *PLoS One*. 2013;8(9):e75810. doi:10.1371/journal.pone.0075810
6. Baker E, Gwernan-Jones R, Britten N, et al. Refining a model of collaborative care for people with a diagnosis of bipolar, schizophrenia or other psychoses in England: a qualitative formative evaluation. *BMC Psychiatry*. 2019;19(1):7. doi:10.1186/s12888-018-1997-z
7. Halcomb EJ, McInnes S, Patterson C, Moxham L. Nurse-delivered interventions for mental health in primary care: a systematic review of randomized controlled trials. *Fam Pract*. 2019;36(1):64-71. doi:10.1093/fampra/cmy101
8. Tawiah AK, Desmeules F, Finucane L, et al. Advanced practice in physiotherapy: a global survey. *Physiotherapy*. 2021;113:168-176. doi:10.1016/j.physio.2021.01.001
9. Vedanayagam M, Buzak M, Reid D, Saywell N. Advanced practice physiotherapists are effective in the management of musculoskeletal disorders: a systematic review of systematic reviews. *Physiotherapy*. 2021;113:116-130. doi:10.1016/j.physio.2021.08.005
10. Kersten P, McPherson K, Lattimer V, George S, Breton A, Ellis B. Physiotherapy extended scope of practice – who is doing what and why? *Physiotherapy*. 2007;93(4):235-242. doi:10.1016/j.physio.2007.02.007
11. Lafrance S, Demont A, Thavorn K, Fernandes J, Santaguida C, Desmeules F. Economic evaluation of advanced practice physiotherapy models of care: a systematic review with meta-analyses. *BMC Health Serv Res*. 2021;21(1):1214. doi:10.1186/s12913-021-07221-6
12. Gérvas J, Pérez Fernández M. Western European best practice in primary healthcare. *European Journal of General Practice*. 2006;12(1):30-33. doi:10.1080/13814780600757187
13. Groenewegen P, Schäfer W, Schellevis F, Boerma W. Kernwaarden van Nederlandse huisartsen in internationaal perspectief. *Huisarts Wet*. 2020;63(7):57-61. doi:10.1007/s12445-020-0760-0
14. Shaw BR, Heywood SE, Page CJ, et al. Advanced musculoskeletal physiotherapy: Barriers and enablers to multi-site implementation. *Musculoskeletal Care*. 2018;16(4):440-449. doi:10.1002/msc.1358
15. Tawiah AK, Borthwick A, Woodhouse L. Advanced Physiotherapy Practice: A qualitative study on the potential challenges and barriers to implementation in Ghana. *Physiother Theory Pract*. 2020;36(2):307-315. doi:10.1080/09593985.2018.1484535
16. Dawson LJ, Ghazi F. The experience of physiotherapy extended scope practitioners in orthopaedic outpatient clinics. *Physiotherapy*. 2004;90(4):210-216. doi:10.1016/j.physio.2004.06.001

17. Morris J, Grimmer K, Ashman B, et al. Principles to guide sustainable implementation of extended-scope-of-practice physiotherapy workforce redesign initiatives in Australia: stakeholder perspectives, barriers, supports, and incentives. *J Multidiscip Healthc*. Published online June 2014:249. doi:10.2147/JMDH.S59473
18. Jones A, Powell T, Watkins D, Kelly D. Realising their potential? Exploring interprofessional perceptions and potential of the advanced practitioner role: a qualitative analysis. *BMJ Open*. 2015;5(12):e009740. doi:10.1136/bmjopen-2015-009740
19. van Raak R. The transition (management) perspective on long-term changes in healthcare. In: Broerse J, Bunders - Aelen J, eds. *Transitions in Health Systems: Dealing with Persistent Problems*. VU University Press; 2010:49-86.
20. Creswell J. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage; 2007.
21. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3(2):77-101. doi:10.1191/1478088706qp063oa
22. Bastiaens F, Barten DJ, Veenhof C. Identifying goals, roles and tasks of extended scope physiotherapy in Dutch primary care- an exploratory, qualitative multi-step study. *BMC Health Serv Res*. 2021;21(1):19. doi:10.1186/s12913-020-05986-w
23. van Dongen J, van Bokhoven L, Goossens W, Daniëls R, van der Weijden T, Beurskens S. Steeds complexere zorg vraagt om teamwerk. *Huisarts Wet*. 2019;62(11):60-63. doi:10.1007/s12445-019-0305-6
24. Zaat J. Huisarts en netwerk. *Huisarts Wet*. 2019;62(1):17-23. doi:10.1007/s12445-018-0390-y
25. van den Dool J, Meijer W. *Zorg Door de Fysiotherapeut; Nivel Zorgregistraties Eerste Lijn, Jaarcijfers 2019 En Trendcijfers 2016-2019*.; 2019. Accessed November 11, 2021. https://www.nivel.nl/nl/nivel-zorgregistraties-eerste-lijn/cijfers-over-zorgverlening/cijfers-zorgverlening-fysiotherapeuten
26. Dutch College of General Practitioners. *NHG Toekomstvisie 2012-2022*.; 2012. Accessed November 11, 2021. https://www.lhv.nl/wp-content/uploads/2021/06/Kerndocument-toekomstvisie.pdf
27. Bijma B. Wat is uw ideale team. *De dokter Dutch General Practitioners Association*. Published online 2020. Accessed November 11, 2021. https://www.lhv.nl/actueel/de-dokter/de-dokter-oktober-2020/wat-is-uw-ideale-team/.
28. Lambregtse C. Nieuwe initiatieven in de huisartsenzorg: bedreiging of kans? *De dokter Dutch General Practitioners Association*. Published online 2021. Accessed November 11, 2021. https://www.lhv.nl/wp-content/uploads/2021/11/DD2106_08-13_Aanbieders.pdf
29. Dutch General Practitioners Association. LHV-afwegingsinstrument Nieuwe aanbieders in de huisartsenzorg. Accessed November 11, 2021. https://www.lhv.nl/product/lhv-afwegingsinstrument-nieuwe-aanbieders-in-de-huisartsenzorg/
30. Jabaaij L, Schers H, van Essen T, Goudswaard L, Schellevis F. Altijd dezelfde huisarts? Een onderzoek naar wens en waardering van patiënten. *Huisarts Wet*. 2006;49(13):934-939. doi:10.1007/BF03084995
31. Lugtenberg M, Zegers-van Schaick JM, Westert GP, Burgers JS. Why don't physicians adhere to guideline recommendations in practice? An analysis of barriers among Dutch general practitioners. *Implementation Science*. 2009;4(1):54. doi:10.1186/1748-5908-4-54
32. Frambach J, van der Vleuten C, Durning S. AM Last Page. Quality criteria in qualitative and quantitative research. *Academic Medicine*. 2013;88(4):552. doi:10.1097/ACM.0b013e31828abf7f


5

An advanced practice physiotherapy model of care in Dutch primary care: an explorative study

Report for the Dutch Advanced Physiotherapy Practitioners Association; 'Nederlandse Vereniging voor Extended Scope Specialisten (In Dutch)

Pellekooren S, Ostelo RWJG, van Tulder MW, Pool-Goudzwaard AL.

Abstract

Background

The deployment of specialists in musculoskeletal conditions, such as Advanced Practitioner Physiotherapists (APP), can be of added value in relieving workload of GPs and maintaining quality of care for patients with musculoskeletal disorders. Despite positive results in the international literature, little is known about the effects of the deployment of APP in Dutch primary care.

Objective

To get a first impression of the deployment of APP in primary care in the Netherlands by identifying APP patient population, health effects, APP-led health care pathways, and cost of both APP care and regular GP care.

Methods

This was an explorative study in which the care for patients with musculoskeletal complaints, traditionally delivered by a GP, was delivered by APPs. Patients were included at four different practices between December 2020 and December 2021. Data were retrieved trough clinical registration forms and web-based questionnaires at baseline, 3-, and 6-months follow-up. Cost for usual care pathways (i.e., GP-led care pathways) were calculated using data from electronic patient records that involved data recorded by GPs as part of clinical patient care.

Results

A total of 109 patients were included and the most common condition was shoulder pain (41%). In more than half of the patients, the duration of complaints was longer than 6 months and 43% of patients had a recurrent complaint for which more than 80% had previously consulted a healthcare provider. A positive trend was seen for almost all health-related outcome measures at 3- and 6-month follow-up. Almost one-third of all APP-led pathways included a referral to a secondary care facility. Of all included patients, 71% were referred to physiotherapy, making it the most common referral. The mean cost per APP-led care pathway was €486 (SD 209) for a follow up period of 6 months and patient-reported health costs averaged €2901 (6824) and €2729 (SD 5715), at 3- and 6-month follow-up, respectively. The mean cost of a GP-led care pathway was €97 (SD 117) per year.

Conclusions

A significant part of patients who consult APP are characterized by long-term recurrent complaints. Relatively high care utilization was found, which may be explained by the inclusion of patients with complex complaints and the stage of development of the role of ESS. Given the limited number of participating APPs and low inclusion rates, results should be interpreted with caution.

Samenvatting

Achtergrond
De inzet van specialisten op het gebied van het houdings- en bewegingsapparaat, zoals Extended Scope Specialisten (ESS), kan van toegevoegde waarde zijn bij het verlichten van de werkdruk van huisartsen en het behouden van kwaliteit van zorg voor patiënten met musculoskeletale aandoeningen. Ondanks positieve resultaten in de internationale literatuur is er weinig bekend over de effecten van de inzet van ESS in de Nederlandse eerstelijnszorg.

Doel
Een eerste indruk krijgen van de inzet van ESS in de eerstelijnszorg in Nederland door de patiëntenpopulatie, gezondheidseffecten, zorgpaden, en kosten van zowel ESS-zorg als reguliere huisartsenzorg in kaart te brengen.

Methode
Dit was een exploratieve studie waarin de zorg voor patiënten met klachten aan het houdings- en bewegingsapparaat, die traditioneel door een huisarts wordt geleverd, werd geleverd door ESS. Tussen december 2020 en december 2021 werden patiënten geïncludeerd in vier verschillende praktijken. Gegevens werden verzameld via klinische registratieformulieren en online vragenlijsten bij aanvang, 3-, en 6-maanden follow-up. Kosten voor reguliere zorgtrajecten (d.w.z. door de huisarts uitgezette zorgtrajecten) werden berekend met behulp van gegevens uit elektronische patiëntendossiers waarin gegevens waren opgenomen die door huisartsen werden geregistreerd als onderdeel van de klinische patiëntenzorg.

Resultaten
In totaal werden 109 patiënten geïncludeerd en de meest voorkomende aandoening was schouderpijn (41%). Bij meer dan de helft van de patiënten was de duur van de klachten langer dan 6 maanden en 43% van de patiënten had een recidiverende klacht waarvoor meer dan 80% eerder een zorgverlener had geraadpleegd. Een positieve trend werd gezien voor bijna alle gezondheidsgerelateerde uitkomstmaten bij 3- en 6-maanden follow-up. Bijna een derde van alle door ESS uitgezette zorgpaden omvatte een verwijzing naar een tweedelijns zorginstelling. Van alle geïncludeerde patiënten werd 71% doorverwezen naar fysiotherapie, waarmee dit de meest voorkomende verwijzing was. De gemiddelde kosten per ESS uitgezet zorgpad bedroegen €486 (SD 209) bij een follow-up periode van 6 maanden en de door de patiënt gerapporteerde gezondheidskosten bedroegen gemiddeld €2901 (6824) en €2729 (SD 5715), bij een follow-up van respectievelijk 3 en 6 maanden. De gemiddelde kosten van een door de huisarts uitgezet zorgtraject bedroegen 97 euro (SD 117) per jaar.

Conclusie
Een aanzienlijk deel van de patiënten die APP raadplegen wordt gekenmerkt door langdurig terugkerende klachten. Bij de identificatie van door APP uitgezette zorgpaden werd een relatief hoog zorggebruik gevonden, wat verklaard kan worden door de inclusie van patiënten met complexe klachten en het stadium van ontwikkeling van de rol van ESS. Gezien het beperkte aantal deelnemende ESS en de lage inclusiecijfers moeten de resultaten met voorzichtigheid worden geïnterpreteerd.

Inleiding

In Nederland groeit de uitdaging om kwalitatief hoogwaardige en toegankelijke zorg te behouden en tegelijkertijd kosten te beheersen. Daarnaast staat de toegang tot de zorg onder druk door de toenemende vraag naar zorg als gevolg van vergrijzing en een stijgend aantal chronisch zieken. Dit maakt dat de werkdruk in de zorgsector hoog is en zorgmedewerkers de sector verlaten.[1]

Een van de groepen zorgmedewerkers die expliciet aangeeft een hoge werkdruk te ervaren zijn huisartsen.[2] Door complexe patiënten, het overnemen van taken van de tweedelijnszorg en meer administratieve taken neemt de werkdruk van huisartsen toe terwijl tegelijkertijd het aantal collega's afneemt.[3] Deze hoge werkdruk leidt onder andere tot suboptimale diagnostiek en een toename van het aantal verwijzingen naar tweedelijnszorg en beeldvormende diagnostiek.[4]

De toename van zorgvragen en de complexiteit van deze vragen maakt ook dat onder huisartsen meer behoefte ontstaat aan ondersteuning en expertise binnen hun praktijk.[5] Dit is bijvoorbeeld terug te zien in de steeds frequentere samenwerking met andere zorgverleners, zoals bijvoorbeeld de inzet van praktijkondersteuners en verpleegkundig specialisten in de zorg voor patiënten met een chronische aandoeningen zoals diabetes en depressie.[6]

Naast de groep patiënten met een chronische aandoening vormen patiënten met klachten aan het houdings- en bewegingsapparaat ook een substantiële groep die een beroep doet op de huisarts.[7]Om de huisarts te ontlasten en bij te dragen aan behouden van kwaliteit van zorg voor deze patiëntengroep zou de inzet van specialist op het gebied van het houdings- en bewegingsapparaat, zoals de Extended Scope Specialist (ESS), een waardevolle aanvulling kunnen zijn.

In Nederland heeft ESS in 2017 zijn intrede gedaan.[8] Dit in navolging van de internationale ontwikkelingen van de Advanced Practitioner Physiotherapists (APPs), die ook wel Extended Scope Practitioners (ESP's) worden genoemd.[9–11] In andere Angelsaksische landen nemen APP's zorg over die traditioneel door een (huis)arts wordt verleend, zoals het stellen en communiceren van een diagnose, triage voor chirurgie of chirurgisch advies, het aanvragen van diagnostische beeldvorming of laboratoriumtests en het voorschrijven/injecteren van medicijnen.[11,12] Uit een recent gepubliceerd overzicht van systematische reviews blijkt dat de inzet van APP's bijdraagt aan de toegankelijkheid van zorg met vergelijkbare gezondheidseffecten, diagnostische accuratesse en patiënttevredenheid.[10] Daarnaast zou de inzet van APP's kunnen leiden tot lagere directe zorgkosten in vergelijking met gebruikelijke zorg.[13]

De inzet van de ESS in de Nederlandse gezondheidszorg zou wellicht een bijdrage kunnen leveren aan verbeteren van de kwaliteit en betaalbaarheid van zorg en lijkt hiermee aan te sluiten op het door het ministerie van VWS geïnitieerde programma, 'De juiste zorg op de juiste plek' (JZOJP),[14] waarbij regionale samenwerkingsverbanden worden gebruikt om goed op elkaar afgestemde zorg, ondersteuning en welzijn te bieden. De ESS betreft twee van de drie pijlers binnen dit programma (duurdere zorg voorkomen en zorg verplaatsen en rondom de mensen organiseren).

Doel en vraagstellingen

Ondanks veelbelovende internationale publicaties over de positieve effecten van ESS in andere landen en de 650 opgeleide ESS in Nederland, is de er weinig bekend over de inzet van ESS in de Nederlandse eerstelijnszorg.

Om een eerste indruk te krijgen van de inzet van ESS in de eerstelijnszorg zijn de volgende onderzoeksvragen geformuleerd:

a. Welke kenmerken hebben patiënten die gezien worden door een ESS?
b. Wat zijn de gevonden gezondheidseffecten bij patiënten met klachten van het houdings- en bewegingsapparaat, waarbij het zorgpad is uitgezet door ESS?
c. Hoe worden zorgpaden door ESS ingericht bij patiënten met klachten van het houdings- en bewegingsapparaat?
d. Wat zijn de kosten van de door ESS uitgezette zorgpaden bij patiënten met klachten aan het houdings- en bewegingsapparaat in vergelijking met de kosten van reguliere zorg, welke bekend zijn uit bestaande literatuur en/of bestaande databases?

Methode

Studie opzet

Om een eerste indruk te krijgen van de patiëntpopulatie, gezondheidseffecten, zorgpaden en kosten voerden we een exploratief onderzoek uit onder ESS-praktijken die gevestigd waren binnen verschillende eerstelijnszorg settingen. De ESS werkzaam binnen deze praktijken waren door een huisarts geautoriseerd om, op basis van taakherschikking, zorg over te nemen voor patiënten met klachten aan het houdings- en bewegingsapparaat.

Deelnemende ESS

Deelnemende ESS waren geregistreerd in het kwaliteitsregister van de beroepsvereniging van ESS (NVES - Nederlandse Vereniging van Extended Scope Specialisten), hadden minimaal vijf jaar werkervaring als manueel- of sportfysiotherapeut en moesten beschikken over een afgeronde postmasteropleiding tot ESS. Daarnaast moest er sprake zijn van een bestaande samenwerking met één of meerdere huisartsen. Deze samenwerking werd voorafgegaan door een training van ESS door een of meerdere huisartsen, waarna een autorisatie verklaring werd afgegeven. Werving van ESS vond plaats door het benaderen van geregistreerde ESS via een wervingsmail en een oproep in de nieuwsbrief van de beroepsvereniging.

Vijf ESS, werkzaam in 4 verschillende praktijken, voldeden aan deze criteria en namen deel aan de studie. In 3 praktijken was 1 ESS werkzaam en in 1 praktijk 2 ESS. Eén praktijk bevond zich in een landelijk gebied en 3 in een stedelijk gebied. Alle praktijken waren toegankelijk via verwijzing door een huisarts, een fysiotherapeut en zelfverwijzing (Directe Toegankelijkheid Fysiotherapie), behalve 1 praktijk die alleen toegankelijk was voor patiënten met schouderklachten op verwijzing van een arts. Alle ESS waren geautoriseerd voor het uitvoeren van een éénmalig diagnostisch consult waarbij zij de volgende taken uitvoerden: het stellen en communiceren van diagnoses, het uitzetten van zorgpaden, en het aanvragen van diagnostische beeldvorming en laboratoriumonderzoek. Eén ESS was ook bevoegd om injectietherapie voor te schrijven.

Deelnemende patiënten

Deelnemende patiënten waren 18 jaar of ouder en consulteerden een ESS vanwege een klacht aan het houdings- en bewegingsapparaat. Patiënten werden uitgesloten van deelname als zij geen Nederlands konden lezen, niet in staat waren online vragenlijsten in te vullen of leden aan ernstige comorbiditeit buiten het musculoskeletale domein. Werving vond plaats via posters in wachtkamers of tijdens het maken van een afspraak

in een huisartsenpraktijk. Vooraf was het streven om 50 patiënten per praktijk te rekruteren.[15] Met een verwachte inclusie van twee patiënten per week per locatie werd de inclusieperiode van zes maanden voldoende geacht. Vanwege achterblijvende inclusie werd de inclusieperiode met zes maanden verlengd. Alle deelnemers gaven voorafgaand aan deelname aan het onderzoek schriftelijk toestemming voor deelname. Dit onderzoek is uitgevoerd in overeenstemming met de Verklaring van Helsinki en ethische goedkeuring is verkregen van de Medisch Ethische Commissie van het VU medisch centrum in Amsterdam; referentienummer 2020.17.

Uitkomsten

Patiëntkenmerken en gezondheidsgerelateerde uitkomsten werden geëvalueerd met behulp van online vragenlijsten. Bij aanvang vulden patiënten vragenlijsten in met betrekking tot demografische gegevens, geloofwaardigheid en verwachtingen ten aanzien van de behandeling (Credibility and Experience Questionnaire),[16,17] zelfredzaamheid (General Self-efficacy Scale),[18] pijnintensiteit (Numerial Rating Scale; NRS),[19] kwaliteit van leven (EQ-5D-5L),[20] fysiek functioneren (PROMIS V1. 2 Fysiek functioneren)[21] en een ziektespecifieke vragenlijst afhankelijk van het type klacht; DASH voor arm- en schouderklachten,[22] KOOS voor knieklachten,[23] HOOS voor heupklachten,[24] ODI voor rugpijn[25,26] NDI voor nekpijn,[27] FAOS voor enkelklachten.[28] Na drie en zes maanden werden pijnintensiteit, kwaliteit van leven, fysiek functioneren en globaal waargenomen effect (GPE-DV)[29] gemeten. Comorbiditeiten werd gemeten met behulp van de Cumulative Illness Rating Scale[30] en door ESS gerapporteerd middels een klinisch registratieformulier.

Door ESS uitgezette zorgpaden werden geëvalueerd aan de hand van een klinisch registratieformulier. De ESS registreerde voor elke patiënt de stappen van het zorgpad, waaronder bijvoorbeeld advies en verwijzing naar andere zorgverleners zoals huisarts, fysiotherapeut, medisch specialist (voor beeldvorming, diagnose of behandeling) of geestelijke gezondheidszorg. Ook was er een optie om eventuele extra stappen te rapporteren die niet waren opgenomen in de vooraf gespecificeerde antwoordopties.

De kosten van zorgpaden werden berekend op basis van de verzamelde gegevens uit de klinische registratieformulieren. In het geval van een verwijzing naar andere zorgverleners, zoals bijvoorbeeld een fysiotherapeut, werd bij het berekenen van de kosten uitgegaan van een gemiddeld aantal behandelingen omdat het daadwerkelijk aantal ontvangen behandelingen niet kon worden vastgesteld op basis van het klinische registratieformulier. Het gemiddelde aantal behandelingen was gebaseerd op nationale gegevens over zorggebruik.[31]

Patiënt gerapporteerde zorgkosten werden verzameld met behulp van een kostenvragenlijst (iMTA).[32] Deze vragenlijst werd afgenomen op drie en zes maanden nadat het eerste consult bij een ESS plaatsvond. De items van deze vragenlijst hadden betrekking op kosten binnen de gezondheidszorg, kosten voor de patiënt en familie, en kosten in andere sectoren.[33]

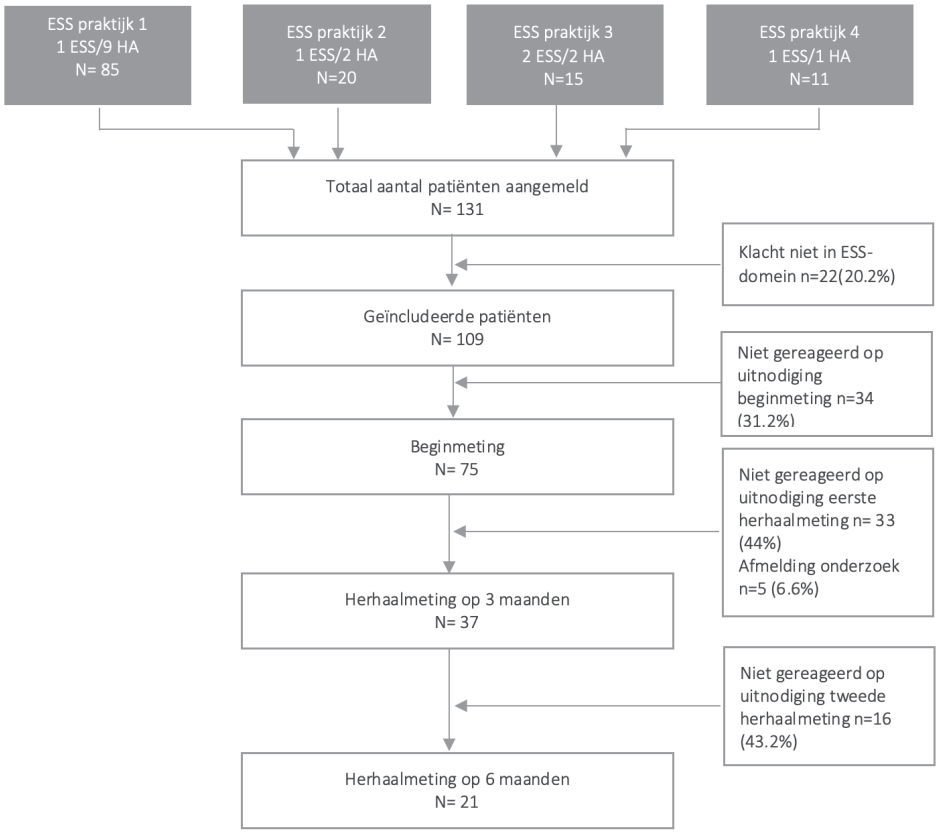
De kosten van huisartsenzorg werden berekend aan de hand gegevens uit elektronische patiëntendossiers van ongeveer 25% van alle patiënten die in huisartsenpraktijken in Nederland worden behandeld. Deze data waren afkomstig uit de database van PHARMO Institute.[34] Data werden geregistreerd door huisartsen als onderdeel van de klinische patiëntenzorg en omvatten prospectief verzamelde anonieme gegevens op patiëntniveau van 403.719 patiënten over consulten, medicatie, verwijzingen, probleemlijsten en tijdschrifteksten. Patiënten van 18 jaar en ouder werden geïncludeerd wanneer in 2018 een International Classification of Primary Care (ICPC) code[35] werd toegekend die een musculoskeletale klacht vertegenwoordigt en patiëntgegevens beschikbaar waren over minimaal één jaar na toekenning van de ICPC-code. Ook bij het berekenen van kosten van huisartsenzorg werd het aantal behandelingen na verwijzing gebaseerd op nationale gegevens over zorggebruik.[31]

De kosten werden berekend met behulp van de Nederlandse richtlijnen voor kostenstudies.[36,37] Indien gegevens niet beschikbaar waren in deze richtlijnen werd het zorgproductenregister van de Nederlandse Zorgautoriteit[38] geraadpleegd. De kosten van medicijnen werden berekend aan de hand van de prijzen van de Koninklijke Nederlandse Maatschappij voor Pharmacie.[39] Het indexjaar voor alle kosten was 2021.

Uitkomsten werden beschreven met behulp van frequenties en percentages voor categorische variabelen, gemiddelden en standaarddeviaties (SD) voor normaal verdeelde continue variabelen, en mediaan en interkwartielafstand (IQR) voor niet-normaal verdeelde continue variabelen en ordinale variabelen. De kosten werden gerapporteerd als gemiddelde en SD en uitgedrukt in euro's.

Resultaten

In totaal werden 131 patiënten geïncludeerd in het onderzoek waarvan meer dan de helft van de patiënten via de huisarts instroomde. Na een eerste screening door ESS bleken 22 patiënten (17%) geen gezondheidsprobleem te hebben binnen de kaders van ESS zorg en werden daarom niet verder meegenomen in het onderzoek. De uitval onder de 109 overgebleven patiënten was 51% en 43% na respectievelijk drie en zes maanden. Meer details over inclusie en uitval zijn weergegeven in Figuur 1.



Figuur 1. Inclusie en follow up

Patiëntenpopulatie

Van de 109 geïncludeerde patiënten vulden 75 patiënten de beginmeting in. Van deze patiënten was 56% vrouw en de gemiddelde leeftijd was 52 jaar (SD 15,4). De gemiddelde score op de comorbiditeiten vragenlijst was 2,4 (SD 2,0) en de gemiddelde score op zelfredzaamheid schaal was 31,6 (SD 4,9). De meest voorkomende aandoening was schouderklachten (41%). Ook wanneer één van de praktijken die alleen schouderklachten behandelde niet meegerekend wordt, dan was schouderpijn nog steeds de meest voorkomende aandoening (36%). Bij meer dan de helft van de patiënten was de duur van de klachten langer dan zes maanden en 43% van de patiënten had een terugkerende klacht waarvoor meer dan 80% van deze patiënten eerder een zorgverlener had geraadpleegd. Meer details over de kenmerken van de patiënten zijn terug te vinden in Tabel 1.

Tabel 1. Patiënt karakteristieken

	ESS-patiënten	**Huisarts patiënten**
	N=75	N=403,719
Leeftijd (gemiddelde, SD)	52.34 (15.42)	52.54 (18.31)
Geslacht (n, %)		
Man	33 (44%)	176.618 (43.7)
Vrouw	42 (56%)	227,101 (56.3)
Body Mass Index (gemiddelde, SD)	26.76 (5.9)	
Regio van de klacht (n, %)		
Wervelkolom totaal	15 (20%)	74,553 (18.5%)
Rug	12 (16%)	
Nek	3 (4%)	
Bovenste extremiteit totaal	(37) 49.3%	109,202 (27.0%)
Schouder	31 (41.3%)	
Arm/pols/hand	6 (8%)	
Onderste extremiteit totaal	(19) 25.4%	113,449 (28.1%)
Heup	8 (10.7%)	
Knie	9 (12%)	
Enkel/voet	2 (2.7%)	
Overige regio	4 (5.3%)	106,515 (26.4%)
Comorbiditeiten score (gemiddelde, SD)	2.39 (1.97)*	2.78 (3.71)**
Self-efficacy Scale (mediaan, IQR)	32.00 (29-35)	
Aanvullende zorgverzekering (n, %)		
Ja	59 (79%)	
Nee	16 (21%)	
Duur van de klachten (n, %)		
<1 maand	6 (8%)	
>1 maand, < 3 maanden	13 (17.3%)	
>3 maanden, < 6 maanden	14 (18.7%)	
>6 maanden	42 (56%)	
CEQ totaal (gemiddelde, SD)	71.6 (20.2)	
Credibility score	32.3 (8.5)	
Expectance score	39.2 (12.2)	
EQ-5D-5L (gemiddelde, SD)		
Utiliteit score	0.712 (0.221)	
VAS score	68.87 (15.82)	

Tabel 1. Continued

	ESS-patiënten N=75	**Huisarts patiënten** N=403,719
Recidiverende klacht (n, %)		
Ja	32 (42.7)	
Nee	43 (57.3)	
Eerder een zorgverlener geconsulteerd (n, %)[a]		
Ja	27 (84.4%)	
Nee	5(15.6%)	
Type zorgverlener dat eerder geconsulteerd is (n, %)[a]		
Huisarts	9 (33.3%)	
Fysiotherapeut	12(44.5%)	
Medisch specialist	6(22.2%)	
Eerder aanvullende diagnostiek ontvangen (n, %)[a]	19 (59.4%)	
Soort aanvullende diagnostiek (n, %)[b]		
Röntgen	10(52.6%)	
Echo	7(36.8%)	
MRI	7(36.8%)	
CT scan	3(15.8%)	
Laboratoriumonderzoek	1(5.3%)	

Self-efficacy Scale: 10 items; range 10-40, hogere scores wijzen op meer zelfredzaamheid

*Comorbidity score: 13 items; range 0-52, hogere scores wijzen op meer en/of ernstigere comorbiditeiten

** Comorbidity score: 0-10, hogere scores wijzen op meer en/of ernstigere comorbiditeiten

EQ-5D-5L utiliteit score: 5 items; range -0.33-1.00, hogere scores wijzen op een beter kwaliteit van leven

CEQ: Credibility and Experience Questionnaire: 11 items; range 11-99, hogere scores wijzen op hogere verwachtingen en geloofwaardigheid

[a] Heeft alleen betrekking op patiënten met recidiverende klachten (n=32)

[b] Heeft alleen betrekking op patiënten die eerder aanvullende diagnostiek hebben ontvangen (n=19)

Gezondheidseffecten

De herhaalmetingen lieten een positief effect zien op pijnintensiteit (0-10 NRS; 3,6 en 2,8), kwaliteit van leven (0-1 EQ-5D-5L; 0,763 en 0,801) en globaal ervaren effect (0-7 schaal; 2,6 en 2,3) op respectievelijk drie en zes maanden. De meting van het lichamelijk functioneren (PROMIS V1. 2 Fysiek functioneren) gaf aan dat patiënten iets meer moeite hadden met lichamelijk functioneren in de loop van de tijd (33,7 na drie maanden en 32,5 na zes maanden). Meer details over de follow-up metingen van gezondheidsgerelateerde uitkomsten staan vermeld in Tabel 2.

Zorgpaden

Van de 109 patiënten hadden 58 (53%) één, 27 (25%) twee en 24 (22%) drie of meer consulten. 77 (72%) patiënten ontvingen zorg binnen een eerstelijnszorg setting, waarbij de meest voorkomende zorgpaden 'advies en fysiotherapie' (42%), 'advies' (11%) en 'advies, injectietherapie en fysiotherapie' (6%) waren. De overige 32 (28%) patiënten ontvingen ook zorg in een tweedelijnszorg setting, waarbij de meest voorkomende zorgpaden 'advies, diagnostische beeldvorming en fysiotherapie' (8%), 'advies, diagnostische beeldvorming en consult medisch specialist' (5%), en 'advies en consultatie medisch specialist' (5%) waren. Van alle geïncludeerde patiënten werd 71% verwezen naar fysiotherapie, waarmee dit de meest voorkomende verwijzing was. Meer details over de verschillende zorgpaden en de kosten per zorgpad staan in Tabel 3.

Kosten van zorgpaden

De gemiddelde kosten van een zorgpad bedroegen € 486 (SD 209). De onderste extremiteit was de regio met de hoogste kosten, namelijk € 506 (SD 223), gevolgd door bovenste extremiteiten en wervelkolom met respectievelijke kosten van € 485 (SD 168) en € 463 (SD 243). Bij het indelen van zorgpaden op basis van setting waren de gemiddelde kosten € 457 (SD 179) en € 510 (SD 219) voor respectievelijk eerstelijns- en tweedelijnszorgpaden. Details over de kosten per eenheid staan in Appendix I.

Patiënt gerapporteerde kosten

De patiënt gerapporteerde kosten bedroegen € 2.901 (SD 6.824) en € 2.729 (SD 5.715) per patiënt, op respectievelijk drie en zes maanden na het eerste consult. Onderste extremiteit was de regio met de hoogste zorgkosten na drie maanden, namelijk € 3.927 (SD 10.078) per patiënt. Na zes maanden was de wervelkolom de regio met de hoogste kosten, namelijk € 3.239 (SD 7.704). Meer details over de kosten staan in Tabel 4. Details over de kosten per eenheid staan in Appendix II.

Kosten huisartsenzorg

De gegevens uit de elektronische huisartsendossiers lieten zien dat de gemiddelde kosten voor een zorgpad, uitgezet binnen de reguliere huisartsenzorg in Nederland, voor alle patiënten met houding en bewegingsklachten € 97 (SD 117) per patiënt per jaar bedroegen. De wervelkolom was de regio met de hoogste kosten, namelijk € 108 (SD 135), gevolgd door onderste extremiteit en bovenste extremiteit met respectievelijke kosten van € 104 (SD120) en € 93 (SD109). Meer details over patiëntkenmerken zijn terug te vinden in Tabel 1. Details over de kosten per eenheid staan vermeld in Appendix II.

Tabel 2. Gezondheidseffecten

Pijn (gemiddelde, SD), (mediaan, IQR)	
NRS gemiddeld ervaren pijn	
NRS ergst ervaren pijn	
Kwaliteit van leven	
EQ-5D-5L utiliteit	
VAS EQ-5D-5L	
Globaal ervaren effect	
GPE	
Fysiek Functioneren (gemiddelde T score, SE)	
PROMIS	
Ziekte specifiek uitkomsten (gemiddelde, SD), (mediaan, IQR)	
ODI	
NDI	
Quick DASH	
HOOS-PS	
KOOS-PS	
FOAS	

NRS: numeric rating scale; range 0-10, hogere scores wijzen op meer pijn
PROMIS V1.2 Physical Function: weergegeven in een T waarde waarbij een algemene populatie een T score heeft van 50. Lagere scores wijzen op lager niveau van functioneren.
Global perceived effect: range 0-7, lagere scores wijzen op een beter effect
EQ-5D-5l utiliteit score; range -0-33 – 1, hogere scores wijzen op een betere kwaliteit van leven.
VAS EQ-5D-5L: range 0-100, hogere scores wijzen op een betere gezondheidstoestand.
ODI: range 0-100, hogere scores wijzen meer moeilijkheden in fysiek functioneren
NDI: range 0-50, hogere scores wijzen meer moeilijkheden in fysiek functioneren

Studie aanvang (n=75)	3 maanden (n=37)	6 maanden (n=21)
5.27 (2.24), 6(4-7)	3.57(2.61), 3(1-6)	2.81(3.06), 1(0-6)
6.28 (2.37), 7(5-8)	4.78 (3.23), 4(2-8)	3.86(3.61), 3(1-8)
0.712(0.221)	0.763 (0.254)	0.801(0.264)
68.87(15.82)	74.05 (18.23)	75.5 (25.84)
-	2.57 (1.28), 2(1-4)	2.29(1.35), 2(1-3)
34.0(1.7)	33.7 (1.7)	32.5(1.7)
Studie aanvang (n=12)	**3 maanden (n=8)**	**6 maanden (n=7)**
44(1.82),	44.5(20.0),	42(20.02),
45(38-49)	41(31.5-53.5)	38(26-58)
Studie aanvang (n=3)	**3 maanden (n=2)**	**6 maanden (n=0)**
21(4)	17.0 (1.41)	NA
21(19-25)	17(16.5-17.5)	
Studie aanvang (n=37)	**3 maanden (n=17)**	**6 maanden (n=10)**
34.66 (19.42)	26.62 (21.06)	20.0 (17.76)
35(20-47.5)	27.5(7.5-32.5)	22.5(1.88-29.38)
Studie aanvang (n=8)	**3 maanden (n=5)**	**6 maanden (n=5)**
10.62(3.7)	5.4(6.43)	5.00 (8.04)
9.5(8-11.7)	2(1-9)	1.50(0.75-5.75)
Studie aanvang (n=9)	**3 maanden (n=5)**	**6 maanden (n=0)**
23 (8.2)	12(7.58)	NA
26(19-28)	11(8-19)	
Studie aanvang (n=2)	**3 maanden (n=0)**	**6 maanden (n=0)**
97.49 (1.19),	NA	NA
97.49(97.07-97.90)		

Quick DASH: range 0-100, hogere scores wijzen meer moeilijkheden in fysiek functioneren
HOOS-PS (ruwe scores): range 0-20, hogere scores wijzen meer moeilijkheden in fysiek functioneren
KOOS-PS (ruwe scores): range 0-28, hogere scores wijzen meer moeilijkheden in fysiek functioneren
FOAS (ADL sub schaal); range 0-100, hogere scores wijzen minder moeilijkheden in fysiek functioneren
NA: Geen patiënten meer geïncludeerd met klachten in deze regio op dit meetmoment
SD: Standaard Deviatie
SE: Standaard Error
IQR: interkwartielafstand

Tabel 3. Kosten van zorgpaden

Consulten ESS		
Type consult		**Kosten per eenheid in euro's**
Eerste consult		68.03
Vervolgconsulten		36.04
Kosten voor eerstelijnszorgpaden (n=77, 70.6%)		
Type zorgpad	**n (%)**	**Additionele kosten in euro's**
Advies	12(11)	0
Fysiotherapie	3(2.8)	443.30
Advies en fysiotherapie	46(42.2)	443.30
Advies, fysiotherapie en injectie	6(5.5)	448.07
Advies, GP en fysiotherapie	1 (0.9)	479.34
Advies, injectie en fysiotherapie	3(2.8)	448.07
Advies, injectie en huisarts	1(0.9)	40.81
Advies, injectie	4(3.7)	4.77
Advies, podotherapie	1(0.9)	279.23
Kosten voor tweedelijnszorgpaden (n=32, 29.4%)		
Type zorgpad	**n (%)**	**Additionele kosten in euro's**
Advies, consult medisch specialist en fysiotherapie (n=1)	2(1.8)	575.98
Fysiotherapie en beeldvorming (n=1)	1(0.9)	594.90
Advies, fysiotherapie en consult medisch specialist (n=1)	1(0.9)	575.98
Advies, fysiotherapie en beeldvorming (n=1)	1(0.9)	594.90
Advies en consult medisch specialist (n=5)	6(4.6)	132.68
Advies en beeldvorming (n=2)	2(1.8)	151.60
Advies, beeldvorming en fysiotherapie (n=9)	9 (8.3)	594.90
Advies, beeldvorming, consult medisch specialist en fysiotherapie (n=1)	2(1.8)	727.58
Advies, beeldvorming en consult medisch specialist (n=5)	4(3.7)	284.28
Advies, fysiotherapie, injectie en consult medisch specialist (n=1)	1(0.9)	580.75
Advies, injectie en consult medisch specialist (n=1)	1(0.9)	137.45
Advies, beeldvorming, injectie en fysiotherapie (n=1)	1(0.9)	599.67
Advies, beeldvorming, injectie	1(0.9)	156.37

Tabel 4. Patiënt gerapporteerde zorgkosten

	3 maanden	6 maanden
Alle regio's gecombineerd	**n= 37**	**n= 21**
Kosten binnen gezondheidszorg (gemiddelde, SD)	368 (795)	301 (603)
Kosten patiënt en familie (gemiddelde, SD)	55 (112)	28 (68)
Kosten in andere sectoren (gemiddelde, SD)	2.478 (6.131)	2.400 (5.162)
Totale kosten (gemiddelde, SD)	2.900 (6.824)	2.729 (5.715)
Wervelkolom	**n=10**	**n=7**
Kosten binnen gezondheidszorg (gemiddelde, SD)	310 (673)	525 (821)
Kosten patiënt en familie (gemiddelde, SD)	44 (86)	35 (59)
Kosten in andere sectoren (gemiddelde, SD)	1.936 (5.727)	2.679 (6.904)
Total kosten (gemiddelde, SD)	2.290 (6389)	3.239 (7.704)
Bovenste extremiteit	**n=17**	**n=10**
Kosten binnen gezondheidszorg (gemiddelde, SD)	259 (417)	228 (532)
Kosten patiënt en familie (gemiddelde, SD)	48 (74)	34 (87)
Kosten in andere sectoren (gemiddelde, SD)	2.715 (6.363)	2.794 (5.083)
Totale kosten (gemiddelde, SD)	3.022 (6.649)	3.056 (5561)
Onderste extremiteit	**n=10**	**n=4**
Kosten binnen gezondheidszorg (gemiddelde, SD)	611 (1.294)	91 (182)
Kosten patiënt en familie (gemiddelde, SD)	78 (180)	0 (0)
Kosten in andere sectoren (gemiddelde, SD)	3238 (8.650)	927 (1.103)
Totale kosten (gemiddelde, SD)	3.927 (10.078)	1.018 (1.181)

Discussie

Deze studie toonde aan dat van de patiënten die een ESS bezoeken een aanzienlijk deel langdurige en recidiverende klachten heeft. Over het algemeen werd een positieve trend gezien op gezondheidsgerelateerde uitkomsten. Meer dan twee derde van de zorgpaden bevatte een verwijzing fysiotherapie en bijna een derde van alle patiënten werd doorverwezen naar een tweedelijns zorginstelling. Daarnaast liet de vergelijking van de zorgkosten van de ESS in deze studie met landelijke kosten van de Pharmo database zien dat de inzet van ESS geassocieerd is met hogere kosten dan reguliere huisartsenzorg met betrekking tot het uitzetten van zorgpaden.

Vergelijking met de literatuur

Een substantieel deel van de naar de ESS verwezen patiënten had een gezondheidsprobleem buiten het domein van ESS zorg. Dit wijst erop dat in de praktijk, zowel onder huisartsen als patiënten, nog onduidelijkheid bestaat over welke patiënten naar een ESS kunnen worden verwezen. Dit zou mogelijk verklaard kunnen worden door onbekendheid met de rol van ESS, maar ook met suboptimale triagecriteria bij de verwijzing naar ESS.[40]

Een aanzienlijk deel van de patiënten verwezen naar de ESS leek ernstiger klachten te hebben dan andere patiëntpopulaties met klachten aan het houdings- en bewegingsapparaat, zoals bijvoorbeeld patiënten die een fysiotherapeut bezoeken.[31] Het is niet duidelijk of ESS geraadpleegd werd door meer complexe patiënten op basis van de behoefte van huisartsen aan meer expertise in musculoskeletale zorg[8], of dat de Covid pandemie leidde tot het meer zien van patiënten met ernstigere aandoeningen, wat ook gold voor huisartsenpraktijken in deze periode.[41]

Het includeren van patiënten met ernstigere klachten kan ook een gedeeltelijke verklaring zijn voor een hoger zorggebruik binnen de door ESS uitgezette zorgpaden dan men zou verwachten. De helft van de geïncludeerde patiënten kreeg twee of meer consulten in plaats van een eenmalig diagnostisch consult en er werden relatief hoge verwijzingspercentages gevonden voor fysiotherapie en tweedelijns zorginstellingen (i.e., medisch specialist en diagnostische beeldvorming). Naast de ernst van de klachten kan het hoge zorggebruik ook te maken hebben met onzekerheid van ESS in hun nieuwe rol, onvoldoende (ervaren) opleiding en onduidelijkheid over het werken in overeenstemming met huisartsenstandaarden.[40] Daarnaast lijkt ESS een ontwikkeling door te maken waarbij deze beroepsgroep zich in toenemende mate richt op complexe en hoog complexe zorg.[8]

De inzet van ESS laat in vergelijking met reguliere huisartsenzorg hogere kosten zien ten aanzien van zorgpaden. Dit is niet in overeenstemming met andere studies[42,43] die lagere kosten vonden in het vergelijken van de inzet van ESS met huisartsenzorg. Deze verschillen komen waarschijnlijk enerzijds voort uit verschillen in organisatie van de zorg en positionering van de huisarts,[44,45] en anderzijds uit verschillen in studieopzet, patiëntkenmerken en overgedragen taken. In de studie van Bornhöft et al [42v]werd een afgebakende patiëntenpopulatie geïncludeerd, waarbij patiënten werden uitgesloten van deelname op het moment dat zij al onder behandeling waren voor de betreffende klacht, er sprake was van chroniciteit, of al een fysiotherapeutische behandeling hadden ondergaan. Ook verschilden de taken die aan ESS werden overgedragen tussen onze studie en deze andere studies.[42,43] In één studie[42] was het onduidelijk of ESS rechtstreeks konden doorverwijzen naar andere zorgverleners, zoals medisch specialisten, en in een andere studie[43] waren ESS alleen bevoegd om te verwijzen naar beeldvormende diagnostiek.

Interpretatie van het gevonden resultaat

Het lage aantal deelnemende praktijken leidt tot een beperkte generaliseerbaarheid van onze resultaten. Ondanks pogingen om meer praktijken te includeren hebben uiteindelijk 4 van de 22 praktijken die in aanmerking kwamen deelgenomen aan het onderzoek. Praktijken die afzagen van deelname konden of wilden niet deelnemen om verschillende redenen, zoals het feit dat zij nog onder supervisie van een huisarts werkten, geen toegang hadden tot elektronische patiëntendossiers van de huisarts, of het ontbreken van een passende vergoeding. De lage inclusie van patiënten leidt tot een beperkte betrouwbaarheid en generaliseerbaarheid van de gevonden resultaten. Tegenvallende inclusie is enerzijds te wijten aan het uitvoeren van de studie tijdens de Covid-pandemie. Vooral patiëntgebonden onderzoek is sterk vertraagd tijdens de Covid-pandemie. Veel bezoeken aan huisartsen en fysiotherapeuten werden uitgesteld. Anderzijds kan de tegenvallende inclusie het gevolg zijn van suboptimale verwijsstromen naar verschillende praktijken. Het onvermogen van ESS om voldoende verwijsstromen tot stand te brengen werd veroorzaakt door verscheidene problemen, zoals bijvoorbeeld het ontbreken van een passende vergoedingsstructuur en te weinig autoriserende huisartsen per ESS. Deze factoren werden ook in de kwalitatieve studie als belemmerende factor geïdentificeerd.[40] De hoge uitval van patiënten gedurende de looptijd van de studie kan mogelijk ook vertekening van de resultaten geven. Het is onduidelijk of dit een over- of onderschatting van de resultaten heeft veroorzaakt, aangezien deze uitval deels te wijten was aan problemen met de distributie van online vragenlijsten (d.w.z. vragenlijsten belandden in de spamfolder of werden verward met andere vragenlijsten die voor kwaliteitsdoeleinden door andere partijen, zoals zorgverzekeraars, werden verstuurd).

Aanbevelingen

Met deze studie is een start gemaakt met het verkennen van de inzet van ESS in de eerstelijnszorg in Nederland. De uitkomsten hiervan kunnen gezien worden als een vertrekpunt voor verder onderzoek naar de inzet van ESS. In vervolgonderzoek zou een raamwerk over het ontwikkelen en evalueren complexe interventies, zoals het Framework for Developing and Evaluating Complex Interventions,[46] richting kunnen geven aan mogelijke vervolgstappen. Dit raamwerk biedt de mogelijkheid om vanuit verschillende perspectieven en methodes vragen rondom complexe interventies te beantwoorden, de meest relevantie uitkomsten vast te stellen in de verschillende ontwikkelfases en onderzoek af te wisselen tussen de verschillende ontwikkelfases.

Conclusie

Een aanzienlijk deel van de patiënten die een ESS consulteert wordt gekenmerkt door langdurig terugkerende klachten en een geschiedenis van diagnostische beeldvorming en eerdere behandeling. Tijdens het in kaart brengen van de door ESS uitgezette zorgpaden werd een relatief hoog zorggebruik gevonden hetgeen mogelijk verklaard zou kunnen worden door het insluiten van patiënten met complexe klachten en de ontwikkelingsfase waarin de rol van ESS zich bevindt. De inzet van ESS laat een positieve trend zien op gezondheidsgerelateerde uitkomsten. Gezien het beperkte aantal deelnemende ESS-praktijken en de lage inclusie moeten de resultaten met voorzichtigheid worden geïnterpreteerd.

Referenties

1. Rijksoverheid. *Integraal Zorgakkoord: "Samen Werken Aan Gezonde Zorg."*; 2022. Accessed November 2, 2022. https://open.overheid.nl/repository/ronl-464b0967c396f0f6cc75069e52d1d1ace9a838a6/1/pdf/Integraal%20Zorg%20Akkoord.pdf
2. Schäfer W, van den Berg M, Groenewegen P. De werkbelasting van huisartsen in internationaal perspectief. *Huisarts Wet*. 2016;59(94):94-101.
3. van der Horst H, Bindels P, Assendelft P, et al. Hoogste tijd voor minder patiënten per huisarts. *Huisarts Wet*. 2018;61(3):30-31. doi:10.1007/s12445-018-0056-9
4. Boekee S, Hoekstra H. *Meer Tijd Voor de Patiënt*.; 2018. Accessed March 4, 2022. https://www.lhv.nl/wp-content/uploads/2021/05/meer_tijd_voor_de_patient_lhv_-_uitkomsten_onderzoek_newcom_15_maart_2018.pdf
5. van Dongen J, van Bokhoven L, Goossens W, Daniëls R, van der Weijden T, Beurskens S. Steeds complexere zorg vraagt om teamwerk. *Huisarts Wet*. 2019;62(11):60-63. doi:10.1007/s12445-019-0305-6
6. Korevaar J, Donker G. Taakverschuiving van huisarts naar verpleegkundige. *Huisarts Wet*. 2019;62(11):92-92. doi:10.1007/s12445-019-0269-6
7. Flinterman L, Groenewegen P, Verheij R. *Zorglandschap En Zorggebruik in Een Veranderende Eerste Lijn*.; 2018.
8. Eskens M, Hallegraeff H. *Beroepscompetentieprofiel Extended Scope Specialisten*.; 2022.
9. Tawiah AK, Desmeules F, Finucane L, et al. Advanced practice in physiotherapy: a global survey. *Physiotherapy*. 2021;113:168-176. doi:10.1016/j.physio.2021.01.001
10. Vedanayagam M, Buzak M, Reid D, Saywell N. Advanced practice physiotherapists are effective in the management of musculoskeletal disorders: a systematic review of systematic reviews. *Physiotherapy*. 2021;113:116-130. doi:10.1016/j.physio.2021.08.005
11. Kersten P, McPherson K, Lattimer V, George S, Breton A, Ellis B. Physiotherapy extended scope of practice – who is doing what and why? *Physiotherapy*. 2007;93(4):235-242. doi:10.1016/j.physio.2007.02.007
12. Marks D, Comans T, Bisset L, Scuffham PA. Substitution of doctors with physiotherapists in the management of common musculoskeletal disorders: a systematic review. *Physiotherapy*. 2017;103(4):341-351. doi:10.1016/j.physio.2016.11.006
13. Lafrance S, Demont A, Thavorn K, Fernandes J, Santaguida C, Desmeules F. Economic evaluation of advanced practice physiotherapy models of care: a systematic review with meta-analyses. *BMC Health Serv Res*. 2021;21(1):1214. doi:10.1186/s12913-021-07221-6
14. Rijksoverheid. *De Juiste Zorg Op de Juiste Plek*.; 2018. Accessed March 24, 2022. https://open.overheid.nl/repository/ronl-7241c365-d8f3-4eb0-befe-6de4dc088f60/1/pdf/de-juiste-zorg-op-de-juiste-plek.pdf
15. Viechtbauer W, Smits L, Kotz D, et al. A simple formula for the calculation of sample size in pilot studies. *J Clin Epidemiol*. 2015;68(11):1375-1379. doi:10.1016/j.jclinepi.2015.04.014
16. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. *J Behav Ther Exp Psychiatry*. 2000;31(2):73-86. doi:10.1016/S0005-7916(00)00012-4
17. Smeets RJEM, Beelen S, Goossens MEJB, Schouten EGW, Knottnerus JA, Vlaeyen JWS. Treatment Expectancy and Credibility Are Associated With the Outcome of Both Physical and Cognitive-behavioral Treatment in Chronic Low Back Pain. *Clin J Pain*. 2008;24(4):305-315. doi:10.1097/AJP.0b013e318164aa75

18. Jerusalem M, Schwarzer R. Self-efficacy as a resource factor in stress appraisal processes. In: Schwarzer R, ed. *Self-Efficacy: Thought Control of Action* . Hemisphere; 1992:195-213.
19. Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA. Studies with pain rating scales. *Ann Rheum Dis*. 1978;37(4):378-381. doi:10.1136/ard.37.4.378
20. EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy (New York)*. 1990;16(3):199-208. doi:10.1016/0168-8510(90)90421-9
21. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol*. 2010;63(11):1179-1194. doi:10.1016/j.jclinepi.2010.04.011
22. Beaton DE, Wright JG, Katz JN. Development of the QuickDASH. *J Bone Joint Surg*. 2005;87(5):1038-1046. doi:10.2106/JBJS.D.02060
23. Perruccio AV, Stefan Lohmander L, Canizares M, et al. The development of a short measure of physical function for knee OA KOOS-Physical Function Shortform (KOOS-PS) – an OARSI/OMERACT initiative. *Osteoarthritis Cartilage*. 2008;16(5):542-550. doi:10.1016/j.joca.2007.12.014
24. Davis AM, Perruccio AV, Canizares M, et al. The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. *Osteoarthritis Cartilage*. 2008;16(5):551-559. doi:10.1016/j.joca.2007.12.016
25. Fairbank JCT, Pynsent PB. The Oswestry Disability Index. *Spine (Phila Pa 1976)*. 2000;25(22):2940-2953. doi:10.1097/00007632-200011150-00017
26. Fairbank J, Couper J, Davies J, O'Brien J. The Oswestry low back pain disability questionnaire. *Physiotherapy*. 1980;66(8):271-273.
27. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. *J Manipulative Physiol Ther*. 1991;14(7):409-415.
28. van den Akker-Scheek I, Seldentuis A, Reininga IH, Stevens M. Reliability and validity of the Dutch version of the foot and ankle outcome score (FAOS). *BMC Musculoskelet Disord*. 2013;14(1):183. doi:10.1186/1471-2474-14-183
29. Hudak PL, Wright JG. The Characteristics of Patient Satisfaction Measures. *Spine (Phila Pa 1976)*. 2000;25(24):3167-3177. doi:10.1097/00007632-200012150-00012
30. Linn BS, Linn MW, Gurel L. Cumulative Illness Rating Scale. *J Am Geriatr Soc*. 1968;16(5):622-626. doi:10.1111/j.1532-5415.1968.tb02103.x
31. Veldkamp R, Kruisselbrink M, Meijer W. *Nivel Zorgregistraties Eerste Lijn – Zorg Door de Fysiotherapeut; Jaarcijfers 2020 En Trendcijfers 2017 – 2020*.; 2022. Accessed November 4, 2022. https://www.nivel.nl/nl/publicatie/zorg-door-de-fysiotherapeut-nivel-zorgregistraties-eerste-lijn-jaarcijfers-2020-en
32. Bouwmans C, Krol M, Severens H, Koopmanschap M, Brouwer W, Roijen LH van. The iMTA Productivity Cost Questionnaire. *Value in Health*. 2015;18(6):753-758. doi:10.1016/j.jval.2015.05.009
33. Drummond M, Sculpher M, O'Brien B, Torrance G, Stoddart G. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd ed. Oxford University Press; 2005.
34. PHARMO institute. Accessed March 22, 2023. https://pharmo.nl
35. Lamberts H, Wood M. *ICPC: International Classification of Primary Care*. Oxford University Press; 1987.
36. Hakkaart-van Roijen L, van der Linden N, Bouwmans C, Kanters T, Tan S. *Kostenhandleiding: Methodologie van Kostenonderzoek En Referentieprijzen Voor Economische Evaluaties in de Gezondheidszorg*.; 2015. Accessed February 14, 2023. https://www.zorginstituutnederland.nl/over-ons/publicaties/publicatie/2016/02/29/richtlijn-voor-het-uitvoeren-van-economische-evaluaties-in-de-gezondheidszorg

37. Kanters TA, Bouwmans CAM, van der Linden N, Tan SS, Hakkaart-van Roijen L. Update of the Dutch manual for costing studies in health care. *PLoS One*. 2017;12(11):e0187477. doi:10.1371/journal.pone.0187477
38. Nederlandse Zorg Autoriteit. Accessed March 4, 2022. https://www.nza.nl
39. Rijksinstituut voor Volksgezondheid en Milieu Ministerie van Volksgezondheid W en Sport. Medicijnkosten. Accessed March 4, 2022. https://www.medicijnkosten.nl
40. Pellekooren S, Donker M, Reijnders E, et al. The introduction of advanced practice physiotherapy within Dutch primary care is a quest for possibilities, added value, and mutual trust: a qualitative study amongst advanced practice physiotherapists and general practitioners. *BMC Health Serv Res*. 2022;22(1):529. doi:10.1186/s12913-022-07906-6
41. Heins M, Hek K, Hooiveld M, Hendriksen J, Korevaar J. *Impact Coronapandemie Op Zorgvraag Bij Huisartsen (Factsheet A). Gevolgen Corona in de Huisartsenpraktijk*.; 2020. Accessed March 4, 2022. https://www.nivel.nl/nl/publicatie/impact-coronapandemie-op-zorgvraag-bij-huisartsen-factsheet-gevolgen-corona-de
42. Bornhöft L, Thorn J, Svensson M, Nordeman L, Eggertsen R, Larsson MEH. More cost-effective management of patients with musculoskeletal disorders in primary care after direct triaging to physiotherapists for initial assessment compared to initial general practitioner assessment. *BMC Musculoskelet Disord*. 2019;20(1):186. doi:10.1186/s12891-019-2553-9
43. Peterson G, PortstrÖm M, Frick J. Extended roles in primary care when physiotherapist-initiated referral to X-ray can save time and reduce costs. *International Journal for Quality in Health Care*. 2021;33(3). doi:10.1093/intqhc/mzab122
44. Gérvas J, Pérez Fernández M. Western European best practice in primary healthcare. *European Journal of General Practice*. 2006;12(1):30-33. doi:10.1080/13814780600757187
45. Groenewegen P, Schäfer W, Schellevis F, Boerma W. Kernwaarden van Nederlandse huisartsen in internationaal perspectief. *Huisarts Wet*. 2020;63(7):57-61. doi:10.1007/s12445-020-0760-0
46. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ*. Published online September 30, 2021:n2061. doi:10.1136/bmj.n2061

Appendix I. Kosteneenheden per zorgpad

Zorgpaden in eerstelijnszorg	
Zorgpad	**n=77**
Advies	12
Fysiotherapie	3
Advies en fysiotherapie	46
Advies, fysiotherapie en injectie	6
Advies, consult huisarts en fysiotherapie	1
Advies, injectie en fysiotherapie	3
Advies, injectie en consult huisarts	1
Advies en injectie	4
Advies en podotherapie	1

Zorgpaden in tweedelijnszorg	
Zorgpad	**n=32**
Advies, consult medisch specialist en fysiotherapie	2
Fysiotherapie en beeldvorming	1
Advies, fysiotherapie en consult medisch specialist	1

			Totale kosten in euro's
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	0
	Fysiotherapie: 36.04 euro per zitting * 12.3 zittingen	443.30	443.30
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	443.30
	Fysiotherapie: 36.04 euro per zitting * 12.3 zittingen	443.30	
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	448.07
	Fysiotherapie: 36.04 euro per zitting * 12.3 zittingen	443.30	
	Injectie: Medicatie kosten	4.77	
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	479.34
	Consult huisarts	36.04	
	Fysiotherapie: 36.04 euro per zitting * 12.3 zittingen	443.30	
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	448.07
	Injectie: Medicatie kosten	4.77	
	Fysiotherapie: 36.04 euro per zitting * 12.3 zittingen	443.30	
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	40.81
	Injectie: Medicatie kosten	4.77	
	Consult huisarts	36.04	
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	4.77
	Injectie: Medicatie kosten	4.77	
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	279.23
	Podotherapy	279.23	

			Totale kosten in euro's
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	575.98
	Consult medisch Specialist (87.36 perifeer + 178 academie)/2	132.68	
	Fysiotherapie: 36.04 euro per zitting * 12.3 zittingen	443.30	
	Fysiotherapie: 36.04 euro per zitting * 12.3 zittingen	443.30	594.90
	Beeldvorming: (echo 93.60 + röntgen 54.23 + CT 135.57 + MRI 323)/4	151.60	
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	575.98
	Fysiotherapie: 36.04 euro per zitting * 12.3 zittingen	443.30	
	Consult medisch Specialist (87.36 perifeer + 178 academie)/2	132.68	

Zorgpaden in tweedelijnszorg	
Zorgpad	**n=32**
Advies, fysiotherapie en beeldvorming	1
Advies en consult medisch specialist	6
Advies en beeldvorming	2
Advies, beeldvorming en fysiotherapie	9
Advies, beeldvorming, consult medisch specialist en fysiotherapie	2
Advies, beeldvorming en consult medisch specialist	4
Advies, fysiotherapie, injectie en consult medisch specialist	1
Advies, injectie en consult medisch specialist	1
Advies, beeldvorming, injectie en fysiotherapie	1
Advies, beeldvorming en injectie	1

Totaal aantal patiënten N= 109
Kosten betreffen additionele zorgkosten naast consulten ESS (deze zijn voor iedere patiënt individueel berekend)

			Totale kosten in euro's
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	594.90
	Fysiotherapie: 36.04 euro per zitting * 12.3 zittingen	443.30	
	Beeldvorming: (echo 93.60 + röntgen 54.23 + CT 135.57 + MRI 323)/4	151.60	
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	132.68
	Consult medisch Specialist (87.36 perifeer + 178 academie)/2	132.68	
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	151.60
	Beeldvorming: (echo 93.60 + röntgen 54.23 + CT 135.57 + MRI 323)/4	151.6	
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	594.90
	Beeldvorming: (echo 93.60 + röntgen 54.23 + CT 135.57 + MRI 323)/4	151.60	
	Fysiotherapie: 36.04 euro per zitting * 12.3 zittingen	443.30	
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	727.58
	Beeldvorming: (echo 93.60 + röntgen 54.23 + CT 135.57 + MRI 323)/4	151.60	
	Consult medisch Specialist (87.36 perifeer + 178 academie)/2	132.68	
	Fysiotherapie: 36.04 euro per zitting * 12.3 zittingen	443.30	
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	284.28
	Beeldvorming: (echo 93.60 + röntgen 54.23 + CT 135.57 + MRI 323)/4	151.60	
	Consult medisch Specialist (87.36 perifeer + 178 academie)/2	132.68	
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	580.75
	Fysiotherapie: 36.04 euro per zitting * 12.3 zittingen	443.30	
	Injectie: Medicatie kosten	4.77	
	Consult medisch Specialist (87.36 perifeer + 178 academie)/2	132.68	
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	137.45
	Injectie: Medicatie kosten	4.77	
	Consult medisch Specialist (87.36 perifeer + 178 academie)/2	132.68	
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	599.67
	Beeldvorming: (echo 93.60 + röntgen 54.23 + CT 135.57 + MRI 323)/4	151.60	
	Injectie: Medicatie kosten	4.77	
	Fysiotherapie: 36.04 euro per zitting * 12.3 zittingen	443.30	
	Advies: Geen additionele kosten (valt binnen het eerste consult)	0	156.37
	Beeldvorming: (echo 93.60 + röntgen 54.23 + CT 135.57 + MRI 323)/4	151.60	
	Injectie: Medicatie kosten	4.77	

Appendix II. Kosten per eenheid patiënt gerapporteerde kosten

Kosten binnen gezondheidszorg	**Kosten per eenheid in euro's**
Eerste consult ESS	68.03
Vervolgconsult ESS	36.04
Consult huisarts	36.04
Telefonisch consult huisarts	18.56
Aan huis consult huisarts	54.60
Zitting fysiotherapie	36.04
Zitting manuele therapie	41.40
Zitting ergotherapie	36.04
Zitting oefentherapie	37.13
Consult diëtist	32.41
Consult psycholoog	100.81
Les fysio sport	9.40
Traject podotherapie	279.23
Thuiszorg palliatief care	79.72
Verwijzing medisch specialist	132.68
Ziekenhuisopname (per dag)	592.44
Revalidatie traject	3247.20
Bezoek spoedeisende hulp	282.82
ECG	55.12
Paracetamol (per tablet)	0.07
NSAIDS (per tablet)	0.08
Opioïden licht (per tablet)	0.04
Kosten patiënt en familie	**Kosten per eenheid in euro's**
Informele zorg (per uur)	15.29
Kosten in andere sectoren	**Kosten per eenheid in euro's**
Productiviteitsverlies (per uur)	
Betaald werk man	41.93
Betaald werk vrouw	34.51
Onbetaald werk	15.29

PART III

Estimating EQ-5D based utility values


6

Can EQ-5D-3L utility values of low back pain patients be validly predicted by the Oswestry Disability Index for use in cost-effectiveness analyses?

Qual Life Res. 2022; doi:10.1007/s11136-022-03082-6

Pellekooren S, Ben ÂJ, Bosmans JE, Ostelo RWJG, van Tulder MW, Maas ET, Huygen FJPM, Oosterhuis T, Apeldoorn AT, van Hooff ML, van Dongen JM.

Abstract

Background

Quality Adjusted Life Years (QALY) are an essential outcome in economic evaluations that assess whether a new intervention is cost-effective (i.e., provides good value for money) compared to an alternative intervention. However, not all economic evaluations among low back pain patients measure quality of life using a preference-based measure which is necessary to calculate Quality Adjusted Life Years (QALY). If a preference-based quality of life measure is missing, utility values may be predicted using other measurement instruments, such as the frequently used Oswestry Disability Index (ODI) which is a condition-specific questionnaire on low back pain (LBP) complaints. However, it is unclear whether this results in valid estimates of the utility values.

Objective

To assess whether regression modelling can be used to predict EQ-5D-3L utility values from the ODI in LBP patients for use in cost effectiveness analysis.

Methods

EQ-5D-3L utility values of LBP patients were estimated using their ODI scores as independent variables using regression analyses, while adjusting for case-mix variables. Six different models were estimated: 1) Ordinary Least Squares (OLS) regression, with total ODI score, 2) OLS, with ODI item scores as continuous variables, 3) OLS, with ODI item scores as ordinal variables, 4) Tobit model, with total ODI score, 5) Tobit model, with ODI item scores as continuous variables, 6) Tobit model, with ODI item scores as ordinal variables. The models' performance was assessed using the explained variance (R^2) and Root Mean Squared Error (RMSE). The potential impact of using predicted instead of observed EQ-5D-3L utility values on cost-effectiveness outcomes was evaluated in two empirical cost effectiveness analysis.

Results

Complete individual patient data of 18,692 low back pain patients were analysed. All models had a more or less similar R^2 (range: 45-52%) and RMSE (range: 0.21-0.22). The two best performing models produced similar probabilities of cost-effectiveness for a range of willingness-to-pay (WTP) values compared to those based on the observed EQ-5D-3L values. For example, the difference in probabilities ranged from 2% to 5% at a WTP of 50,000 €/QALY gained.

Conclusions

Results suggest that the ODI can be validly used to predict low back pain patients' EQ-5D-3L utility values and QALYs for use in cost-effectiveness analyses.

Introduction

Low Back Pain (LBP) has an estimated incidence of 250 million people worldwide and is characterized by a high burden of disease.[1] Patients with LBP typically experience difficulties in different aspects of health-related quality of life, such as their daily functioning, social participation,[2,3] and working ability.[4,5] These difficulties may affect patients' health-related quality of life considerably,[3,6] and have significant impact on healthcare and societal cost.[7,8] As limited (healthcare) resources are available, decision-makers are not only interested in the effectiveness of LBP treatments recommended in international guidelines, but also in their cost-effectiveness compared to alternative treatments.

Cost effectiveness analysis provide insight into relative cost-effectiveness of treatments by comparing their incremental costs to their incremental effects.[9] These effects are often expressed in Quality-Adjusted Life-Years (QALYs), which combine both the quality and quantity of life into a single outcome.[10] For estimating QALYs, health-related quality of life is typically measured using preference-based quality-of-life measures. Health states obtained from these measures can be converted into utility values, which represent the preferences of the general population of a country for given health states.[11] In many countries, it is recommended to estimate utility values using the EuroQol five-dimension questionnaire (EQ-5D) and national tariffs to account for the fact that health state preferences differ across countries.[12–14] Unfortunately, EQ-5D data are not always available in clinical trials,[15] as higher priority is sometimes given to condition-specific measures that assess more clinically relevant outcomes.[16]

When utility values are missing, QALYs cannot be calculated. However, information about incremental cost per QALY gained is typically required by healthcare decision-makers, particularly at the national level.[12,13] In the absence of the EQ-5D or another generic preference-based quality-of-life measure, a condition-specific measure might be used to predict utility values.[17] In LBP, one of the most frequently used condition-specific measures is the Oswestry Disability Index (ODI).[18] The ODI measures limitations of a patient's performance,[19] and is recommended in the core outcome set for clinical trials in nonspecific LBP[20] and management of LBP.[21]

A previous study assessed the predictive ability of the ODI in estimating utility values from the EQ-5D-3L by using data from 14,544 patients with lumbar degenerative pathology treated in a tertiary spine centre.[22] Linear regression analysis was performed to predict the patients' EQ-5D utility values based on their ODI total or individual item scores and patients reported severity of back and leg pain. Based on a root mean squared error

(RMSE) of 0.14, authors concluded that it is not possible to estimate EQ-5D-3L utility values based on the ODI. However, given the bounded nature of EQ-5D data as well as the possible existence of other contextual factors that influence health-related quality of life in LBP, it is likely that the models' performance might be improved by using a Tobit model to account for possible ceiling effects. The model's performance might also be improved by including a wider variety of LBP patients treated in various settings, while adjusting for more case-mix variables. Moreover, the authors only based their conclusions on the models' RMSE without assessing the impact of using predicted utility scores in cost-effectiveness. Therefore, this study aimed to assess the feasibility of using different regression models to predict EQ-5D-3L utility values in LBP patients based on the ODI in cost-effectiveness analyses while adjusting for a broad range of case mix characteristics.

Method

Source of data

Individual patient data included in this study originated from four previously conducted prospective studies; i.e., the minimal interventional treatments (MINT) study, the rehabilitation after lumbar disc surgery (REALISE) study, the Nijmegen Decision Tool study, and a study evaluating a treatment-based classification system.[23–32] These studies were conducted among sub-acute and chronic LBP patients treated in primary care, secondary care, and/or tertiary care. For all patients, various sociodemographic variables were assessed at baseline, and both the ODI and EQ-5D-3L utility values were assessed at baseline and at one or more follow-up moments. In total, 21,500 patients were included in these studies. For developing the models, only baseline data were used in the present study, because the proportion of participants with missing data was low at baseline (i.e., <5%), thereby preventing the need for imputation of missing values. To assess the final models' performance in a trial-based cost effectiveness analysis setting, baseline as well as follow-up data were used of the MINT study,[23-25] and the treatment-based classification system study.[29,30]

The MINT study, [23-25] the REALISE study,[31,32] and the treatment-based classification system study[29,30] obtained ethical approval from the Medical Ethics Committee of the Erasmus Medical Centre Rotterdam or Medical Ethics Committee of the VU University Medical Centre in Amsterdam. For the Nijmegen Decision Tool study,[26-28] ethical approval was not required, because the "*Dutch Act on Medical Research involving Human Subjects*" does not apply to screening questionnaires that are part of routine practice. More detailed information on the design and study population of the different studies is provided in Appendix I.

Utility values

Utility values were based on the EQ-5D-3L, which is a generic preference-based measure that asks participants to describe their health state on five health dimensions (i.e., mobility, self-care, usual activities, pain/ discomfort, and anxiety/depression) using three severity levels (i.e., no problems, moderate problems, and severe problems).[33] The participants' EQ-5D-3L health states were converted into utility values using the Dutch tariff.[34] Utility values are presented on a continuous scale that is anchored at 1 (indicating full health) to 0 (indicating a state as bad as being dead). Negative values may also occur, which represent health states that are regarded as worse than a state that is as bad as being dead.[10] Dutch EQ-5D-3L utility values can range between -0.33 and 1.

Oswestry Disability Index

The ODI measures the limitations of a patient's performance compared with that of a fit person and consists of ten items assessing various aspects of daily living (e.g., lifting, walking, and travelling). Each item is scored on a six-point scale, ranging from 0 to 5. The overall ODI score was estimated by summing the values of all individual items, subsequently dividing this score by the total possible score, and multiplying this score by 100. The total score ranges from 0 to 100%, with higher scores indicate higher level of disability.[19,35] For this study, the "sex life" (item 8) was not included, as this item is frequently omitted in applied studies as well.[36-38] Including this item would have hampered the generalization of the results to a large number of LBP studies. The cross cultural adapted Dutch language version of the ODI version 2.1a was used in all studies included.[39]

Predictors

The following case-mix variables were included; age (years), gender (male/female), education level (low/moderate/high), living together with a partner (yes/no), type of LBP (sub-acute/chronic), setting (primary care/secondary care/ tertiary care), and back pain (Numeric Rating Scale (NRS: 0-10) Pain score: low 0-3, moderate 4-6, and severe 7-10).[40] Given error proneness of overly detailed models and benefits of ease of use, NRS scores were categorised using cut-off points from an earlier conducted study, which categorised NRS pain scores based on pain-related interference with functioning in patients with chronic musculoskeletal pain.[41] These variables were included, because they were expected to increase the predictive value of the models[42-47] and to be measured in most applied studies, thereby increasing applicability of the models.

Statistical analysis

Baseline characteristics were described using frequencies and percentages for categorical variables and means and standard deviations for continuous variables. Prior to the development of the models, linearity, and additivity assumptions (i.e.,

normally distributed residuals, homoscedasticity, influential cases, and outliers) were assessed using diagnostic plots (i.e., scatterplot, density plot, and boxplots), and diagnostic tests (e.g., Grubbs test). Pearson's correlation coefficient was used to assess the strength of the linear relationship between the patients' EQ-5D-3L based utility values and ODI total scores. To assess the agreement between the EQ-5D-3L and the ODI the Intra Class Correlation (ICC) was calculated using a two-way random effects model.

Model development and variable selection

Models were developed using two regression techniques; i.e., Ordinary Least Squares (OLS) regression and Tobit regression (i.e. censored or truncated regression). OLS regression was included, because it is still one of the most frequently used linear modelling techniques. OLS regression is used to estimate the strength of the association between a continuous outcome variable and one or more independent variables.[48] OLS, however, does not take into account the bounded nature of utility values which can be accounted for in a Tobit regression.[49] This model can estimate linear relationships between variables, where the range of the dependent variable is constrained. This is done using a so-called latent variable that accounts for the fact that the true independent variable is – in our case – bounded at 1. Hereby, biased and inconsistent estimates, that may occur when using OLS regression, may be prevented.[50]

For both the OLS and Tobit model, three different regression models were developed; 1) including the overall ODI score as independent variable, 2) using all nine ODI items scores as independent variables and assuming them to be continuous, and 3) using all nine ODI items scores as independent variables and assuming them to be ordered. This resulted in six different models: 1) OLS, with the total ODI score, 2) OLS, with the ODI item scores as continuous variables, 3) OLS, with the ODI item scores as ordinal variables, 4) Tobit model, with the total ODI score, 5) Tobit model, with the ODI item scores as continuous variables, 6) Tobit model, with the ODI item scores as ordinal variables. To assess which variables increased the predictive value of the models, a bi-directional stepwise selection procedure,[51] using Akaike Information Criterion (i.e., the trade-off between the goodness of fit of the model and the simplicity of the model),[52] with a 5% significance level was used. Stepwise selection combines the elements of forward and backward selection by sequentially adding variables, based on the most contributing predictors, and omitting variables that no longer provide an improvement in the model fit after adding a new variable to the model. Final models only included case-mix variables that increased the predictive value.

Model performance and internal validation

The original dataset was split into a training sample (70%), and a validation sample (30%) using the 'create Data Partition' function in R. This function creates a balanced split of the data by performing a stratified random split of the data based on the mean of the dependent variable, which leads to a comparable mean EQ-5D-3L utility value in both the training and validation dataset. After developing the models in the training sample, their performance was assessed in the validation sample using the RMSE (i.e., the absolute fit of the model) and the adjusted R^2 (i.e., the relative fit of the model). The minimal important difference (MID) of the EQ-5D-3L was used to determine an acceptable RSME, which was set at a cut of point of 0.03.[53] A correlation of 0.5 or higher (i.e., a relatively moderate correlation as the R squared indicates that about half of the variance of the utility values is explained by the ODI) was considered sufficient for performing regression analysis. Recommended models were selected based on parsimony, which is the trade-off between between simplicity of the model (i.e., low AIC) and explanatory predictive power (i.e., high R^2). To assess agreement between the actual and estimated EQ-5D-3L based utility values a Bland Altman analysis was performed for all models.

Sensitivity Analyses

In addition to the main analysis, three sensitivity analyses (SA) were performed. In the first sensitivity analysis (SA1) the variable mental health status was added to the case-mix variables (SA1). SA1 was only performed on a sub-set of the data, as only one of the four datasets (i.e., the MINT study[23-25]) assessed mental health using the Four Dimensional Symptom Questionnaire (4DSQ),[54] and only part of the sample (n=4,123) completed this questionnaire. The 4DSQ assesses four different aspects of mental health (i.e., distress, depression, anxiety, and somatisation), all of which were included in the models as a separate variable. In SA2, the variable living with a partner was omitted. In SA3 the patients' EQ-5D-3L utility values were converted to EQ-5D-5L utility values using the reverse crosswalk (SA3).[55] Reversed cross walk values make it possible to link EQ-5D-3L responses to EQ-5D-5L value sets and can be used when 5L values are wanted but only 3L data is available.[55,56] The 5-level EQ-5D version is an adapted version of the EQ-5D-3L, which is known to be more sensitive and has less ceiling effects, including through changing the number of levels of perceived problems per dimension from 3 to 5.[57]

Cost effectiveness analysis

To assess the models' impact on cost-effectiveness outcomes, complete cases from two randomized controlled trials were used, i.e., empirical dataset 1 (n=68; Apeldoorn et al[29,30]) and empirical dataset 2 (n=424; Maas et al[23-25]). In both studies, QALYs were estimated based on both the actual EQ-5D-3L scores (i.e., actual QALY values) and

based on the patients' ODI scores (i.e., predicted QALY values). Agreement between the actual and estimated EQ-5D-3L based utility values was assessed by performing a Bland Altman analysis for each of the empirical datasets.

Then, full trial-based cost effectiveness analyses were conducted for each of the six models as well as the patients' actual QALY values (i.e., QALYs based on the measured EQ-5D-3L scores). For each trial-based cost effectiveness analysis, mean differences in costs and QALYs between treatment groups were estimated using seemingly unrelated regression analyses. Incremental Cost-Effectiveness Ratios (ICERs) were calculated by dividing the difference in costs by the difference in effects. Uncertainty around cost and QALY differences was estimated using bootstrapping. The percentage of bootstrapped cost-effect pairs was reported per quadrant of the Cost-Effectiveness Plane (i.e., north-east, south-east, north-west, and south-west). Subsequently, Cost-Acceptability Curves (CEACs) were plotted. CEACs indicate an intervention's probability of cost-effectiveness compared to control for a range of willingness-to-pay (WTP) values (i.e., thresholds of 0, 30,000 euro and 50,000). These probabilities were assessed on their decision sensitivity (i.e., how sensitive the conclusion of a cost effectiveness analysis is to using a particular statistical method).[58] Analyses were performed in R software, version 3.4.0.

Results

Participants

Out of the individual patient data that included 21,500 patients, 18,692 complete cases were included for analysis. These patients had sub-acute (n=3248) or chronic LBP (n=15,444). The mean age of the patients was 53.9 years (SD=14.7, range 18.1-91.9) and 61% of the sample was female. The patients' mean ODI score at baseline was 41.23 (SD=15.4, range 0-100) and their mean baseline EQ-5D-3L based utility value was 0.46 (SD=0.29, range -0.3290-1.00). More details on the patients' characteristics are shown in Table 1.

Variables included and model performance

The diagnostic plots showed a linear relationship between EQ-5D-3L based utility values and the ODI, and homogeneity of variance of the residuals. Even though the patients' baseline EQ-5D-3L based utility values followed a bimodal distribution, the corresponding residuals were normally distributed. Hence, the normality of residuals assumption of linear regression was met. No outliers or influential cases were identified. Pearson's correlation coefficient between the patients' baseline EQ-5D-3L utility values and ODI total score was 0.63. The ICC showed an agreement of 0.23 between individual ODI items and EQ-5D-3L items.

Table 1. Baseline characteristics of patients included

Characteristic	**n=18,692**
Age (mean (SD), range)	53.9 (14.7), 18.1-91.9
Gender; female (n, %)	11,345 (60.7)
Education (n, %)	
Low (no education, primary level education, lower vocational and lower secondary education)	5,398 (28.9)
Moderate (higher secondary education or undergraduate)	9,078 (48.6)
High (tertiary, university level, postgraduate)	4,216 (22.6)
Living with a partner (n, %)	14,085 (75.4)
Type of LBP (n, %)	
Subacute (< 3 months)	3,248 (17.4)
Chronic (> 3 months)	15,444 (82.6)
Post-surgery (n, %)	1,587 (8.5)
Setting (n, %)	
Primary care (i.e., physiotherapy clinics)	150 (0.8)
Secondary care (i.e., pain clinics)	4,123 (22.1)
Tertiary care (i.e., hospital)	14,419 (77.1)
NRS Pain (mean (SD))	6.99 (1.9)
Utility score (mean (SD), range)	0.467 (0.299), -0.3290-1.00
ODI score[a] (mean (SD), range)	41.23 (15.4), 0-100
ODI scores per item	
ODI 1 mean (SD)/ median (IQR)	2.66 (0.93) / 3 (2-4)
ODI 2 mean (SD)/ median (IQR)	1.11 (1.04) / 1 (0-2)
ODI 3 mean (SD)/ median (IQR)	2.78 (1.32) / 3 (2-4)
ODI 4 mean (SD)/ median (IQR)	1.44 (1.22) / 1 (0-2)
ODI 5 mean (SD)/ median (IQR)	2.11 (1.09) / 2 (1-3)
ODI 6 mean (SD)/ median (IQR)	2.85 (1.29) /3 (2-4)
ODI 7 mean (SD)/ median (IQR)	1.49 (1.09) / 1 (0-2)
ODI 9 mean (SD)/ median (IQR)	2.14 (1.20) / 2 (1-3)
ODI 10 mean (SD)/ median (IQR)	1.98 (1.32) / 2 (1-3)

[a] excluding item 8 sex life

LBP = Low Back Pain; NRS= Numeric Rating Scale (range 0-10); Utility (range -0.33 to 1); ODI = Oswestry Disability Scale (range 0-100); ODI individual item (range 0-5) ; SD= Standard Deviation IQR= Inter Quartile Range

Table 2. Performance measures in the training set

Model 1: OLS with ODI total scores
Model 2: OLS with ODI individual item total scores continuous
Model 3: OLS with ODI individual item total scores ordered
Model 4: Tobit with ODI total scores
Model 5: Tobit with ODI individual item total scores continuous
Model 6 Tobit with individual item total scores ordered

OLS: Ordinary Least Squares Regression, ODI: Oswestery Disability Index, R^2: proportion of variance for the dependent variable, RMSE: Root Mean Squared Error, AIC: Akaike Information Criteria

An overview of the independent variables that were included in the final models, as well as their respective regression coefficients, can be found in Appendix II. The case-mix variables age, gender, education, partner and NRS were included in all models, whereas type of LBP was not included in any of the models. The variable setting was included in all models except for model 1 (i.e., OLS with ODI total scores). In the models using Tobit regression, 74 of the 13,087 observations in the training set were right censored.

The performance of the different models was more or less the same, with explained variances ranging from 45% to 51% and RMSEs ranging from 0.21 to 0.22. Based on parsimony of the models, models 2 and 5 seem most appropriate to use. More details on the performance of the different models are shown in Table 2.

The mean difference between estimated and actual utility values for model 2 was -0.068 (95%CI -0.495, 0.359), and for model 5 -0.086 (95%CI -0.512, 0.341). Bland Altman plots of models 2 and 5 are shown in Figure 1. The plots for other all models are presented in Appendix III.

Sensitivity analysis

Adding mental health variable(s) to the models resulted in an increase of the explained variance of 2-4%, whereas the RMSE remained similar. Omission of the variable 'living with a partner' (SA2) did not change the models' performance. Using the patients' reversed cross-walked EQ-5D-5L utility values (SA3) improved the models' explained variance by 3-4%, and the RMSE reduced with 0.06-0.07. More details on the results of the sensitivity analyses are provided in Appendix IV.

Performance in the training set (n=13,087)			Performance in validation set (n=5,605)		
R^2	RMSE	AIC	R^2	RMSE	AIC
0.45	0.22	-2326.48	0.46	0.22	-1083.26
0.50	0.21	-3423.24	0.50	0.21	-1513.73
0.51	0.21	-3769.51	0.52	0.21	-1638.09
0.45	0.22	-2061.91	0.46	0.22	-951.61
0.50	0.21	-3164.37	0.50	0.21	-1385.32
0.51	0.21	-3474.88	0.52	0.21	-1494.06

Results cost effectiveness analysis

The mean difference between estimated and actual utility values for empirical dataset 1 model 2 was -0.039 (95%CI -0.075, -0.002), and for model 5 -0.057 (95%CI -0.097, -0.018). The mean difference between estimated and actual utility values for empirical dataset 2 model 2 was 0.295 (95%CI 0.246, 0.344), and for model 5 the mean difference was 0.294 (95%CI 0.248, 0.341). Bland Altman plots of models 2 and 5 for both empirical datasets are shown in Figure 2. The plots for other all models are presented in Appendix V.

In both empirical datasets, the difference between the predicted and actual differences in QALYs was small for the two most parsimonious models (i.e., models 2 and 5:$\Delta \leq 0.004$) and the distributions of cost-effect pairs across the four quadrants of the cost-effectiveness plane were comparable. The cost-effectiveness acceptability curves based on both predicted and actual QALY values were also similar. The predicted probability of an intervention being cost effective at a willingness to pay of 50,000 was slightly higher in both models than the actual probabilities (i.e., 2-5% in model 2, and 3-5% in model 5). More details on the cost-effectiveness outcomes for all models in both empirical studies are shown in Table 3 and Figure 3.

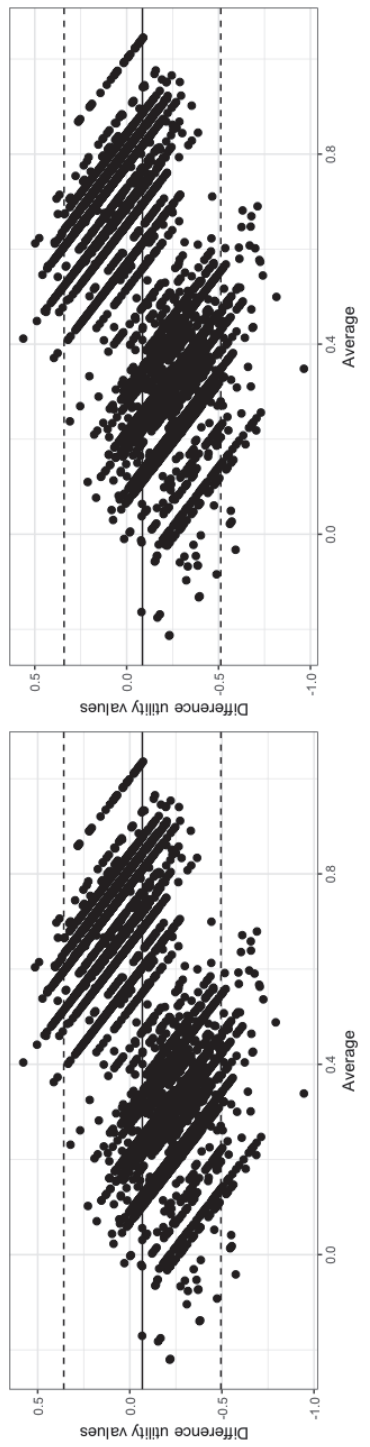


Figure 1. Bland Altman plot model 2 and 5 validation dataset

X-axis: Average measurement of the estimated and actual utility values, Y-axis: Difference in measurements between the two instruments. Solid line: Average difference in measurements between the estimated and actual utility values, Dashed lines: 95% confidence interval limits for the average difference.

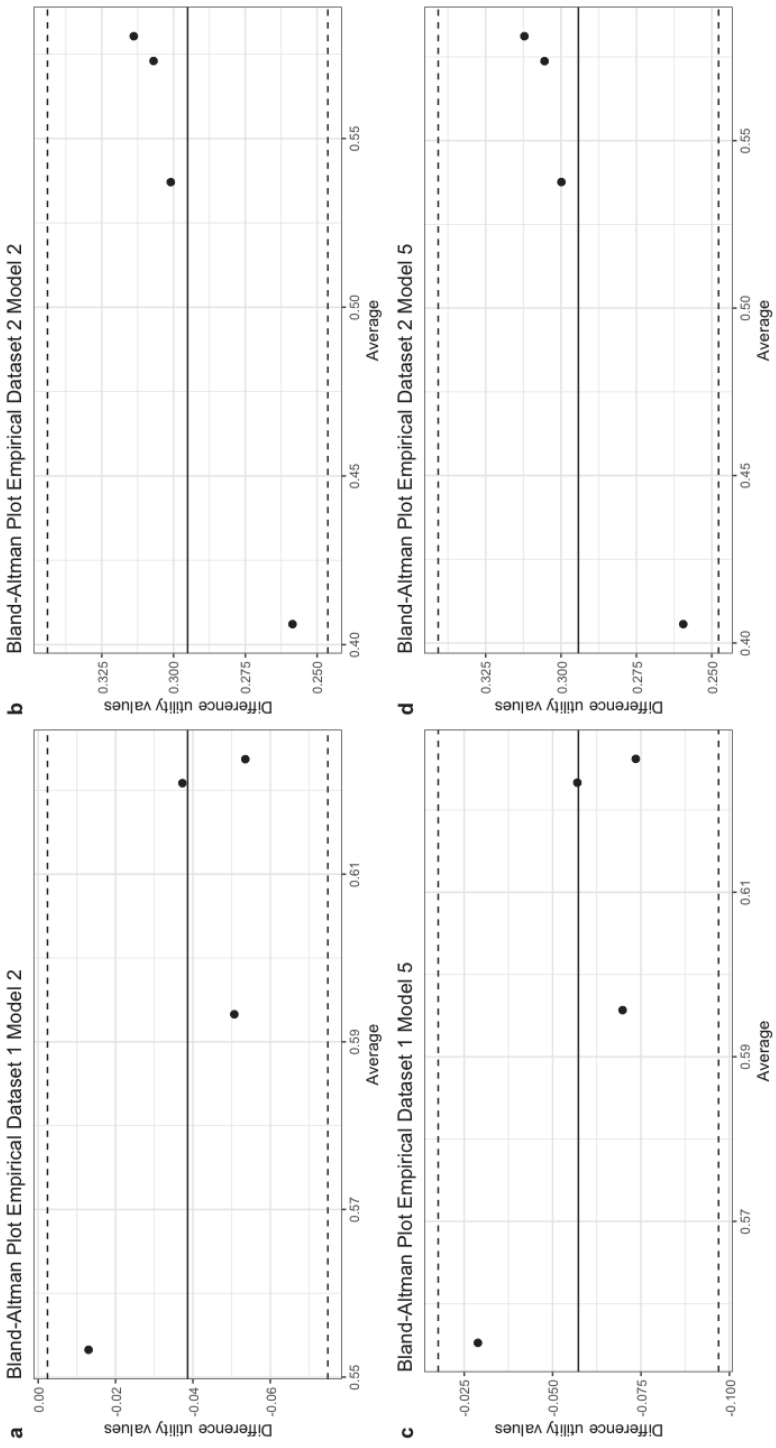


Figure 2. Bland Altman plot model 2 and 5 empirical datasets

X-axis: Average measurement of the estimated and actual utility values, Y-axis: Difference in measurements between the two instruments. Solid line: Average difference in measurements between the estimated and actual utility values, Dashed lines: 95% confidence interval limits for the average difference.

Table 3. Cost-effectiveness outcomes for an intervention in comparison with usual care by predictive models

Predictive models	DE (95% CI)	DC (95% CI)	ICER
Empirical dataset 1 [28, 29] n=86			
Actual values	-0.041 (-0.091; 0.009)	-110 (-1761; 1283)	2697
Model 1	-0.035 (-0.094; 0.021)	-110 (-1761; 1283)	3091
Model 2	**-0.043 (-0.106; 0.015)**	**-110 (-1761; 1283)**	**2559**
Model 3	-0.027 (-0.081; 0.018)	-110 (-1761; 1283)	4068
Model 4	-0.036 (-0.095; 0.021)	-110 (-1761; 1283)	3058
Model 5	**-0.044 (-0.107; 0.015)**	**-110 (-1761; 1283)**	**2514**
Model 6	-0.027 (-0.080; 0.021)	-110 (-1761; 1283)	4084
Empirical dataset 2 [22-24] n=424			
Actual values	-0.004 (-0.034; 0.027)	1576 (596; 2575)	-371566
Model 1	-0.007 (-0.037; 0.023)	1576 (596; 2575)	-226441
Model 2	**0.0002 (-0.030; 0.029)**	**1576 (596; 2575)**	**6670132**
Model 3	-0.001 (-0.026; 0.024)	1576 (596; 2575)	-2099247
Model 4	-0.007 (-0.037; 0.024)	1576 (596; 2575)	-224080
Model 5	**0.0003 (-0.030; 0.030)**	**1576 (596; 2575)**	**5105447**
Model 6	-0.001 (-0.027; 0.026)	1576 (596; 2575)	-2417793

Recommended models are presented as bold text

N = number of observations in the analysis; DC= difference in costs; 95% CI = 95% confidence interval; DE= difference in effects; ICER = Incremental Cost-Effectiveness Ratio; NE = northeast; SE = southeast; SW = southwest; NW = northwest;

Cost-effectiveness plane				Cost-effectiveness acceptability curve			
NE	**SE**	**SW**	**NW**	**P_{CE} (0)**	**P_{CE} (10,000)**	**P_{CE} (30,000)**	**P_{CE} (50,000)**
2%	4%	51%	42%	0.55	0.36	0.16	0.11
1%	10%	45%	44%	0.55	0.39	0.25	0.20
1%	**7%**	**48%**	**44%**	**0.55**	**0.36**	**0.21**	**0.16**
1%	13%	43%	43%	0.55	0.42	0.30	0.24
1%	10%	45%	44%	0.55	0.39	0.25	0.20
1%	**7%**	**48%**	**44%**	**0.55**	**0.36**	**0.21**	**0.16**
2%	13%	42%	43%	0.55	0.42	0.30	0.25
38%	0%	0%	62%	0.001	0.002	0.017	0.048
32%	0%	0%	68%	0.001	0.002	0.014	0.037
51%	**0%**	**0%**	**49%**	**0.001**	**0.003**	**0.025**	**0.070**
48%	0%	0%	52%	0.001	0.002	0.015	0.028
32%	0%	0%	67%	0.001	0.002	0.014	0.038
51%	**0%**	**0%**	**49%**	**0.001**	**0.003**	**0.025**	**0.073**
48%	0%	0%	51%	0.001	0.002	0.018	0.053

P_{CE} (0) = probability that the intervention is cost-effective as compared to usual care with a threshold of 0; P_{CE} ( ) = probability that the intervention is cost-effective as compared to usual care with willingness-to-pay thresholds of 0, 10,000, 30,000, and 50,000 Euros.

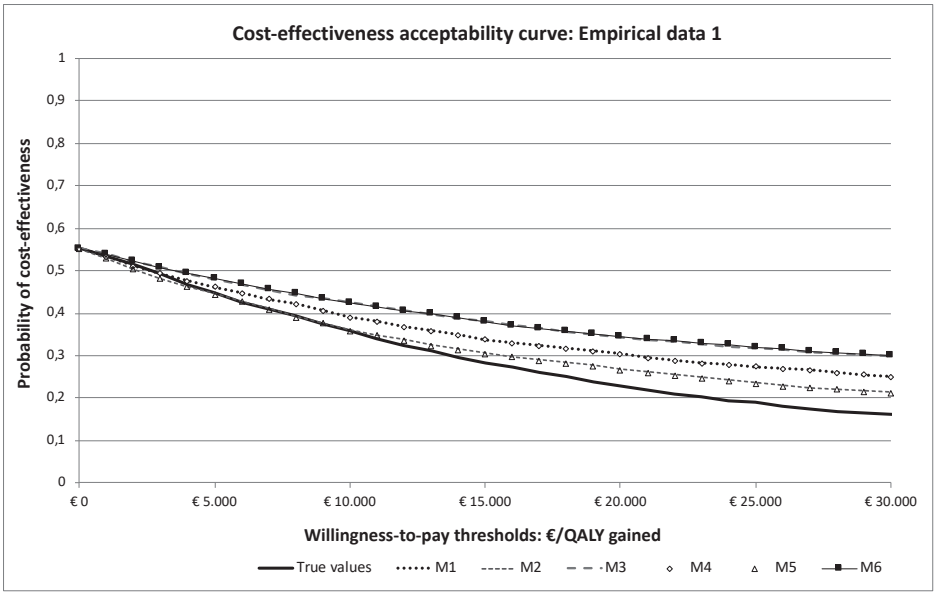


Figure 3. Cost-effectiveness acceptability curves empirical dataset 1

M1= Model 1; M2= Model 2; M3= Model 3; M4= model 4; M5= Model 5; M6= model 6

Discussion

Main findings

There were no large differences in the models' performance between OLS and Tobit regression, nor between using the patients' total ODI scores and ODI individual item scores. The explained variance of the developed models ranged from 45% to 51%, and the RMSE ranged from 0.21 to 0.22. Models 2 and 5 are recommended based on the best fit and parsimony. The models' relatively low absolute fit (RMSE) indicates that they are not suitable for estimating utility values for individual patients. Nonetheless, they can be used to predict differences in LBP patients' EQ-5D-3L utility values and QALY's, as the systematic bias in mean scores does not affect the differences between the groups. Cost-effectiveness outcomes of models 2 and 5 based on predicted and actual values were similar. These findings enable researchers to perform a cost effectiveness analysis with QALYs as the outcome measure, even if EQ-5D-3L data are missing.

Comparison with literature

Our findings regarding the performance measures are more or less in line with the previous study by Carreon et al,[22] who aimed to predict individual LBP patients' EQ-5D-3L utility values based on their ODI scores. Their model performed slightly better

in terms of its explained variance (i.e., R^2 was 61%) and its absolute fit (i.e., RMSE is 0.149), which is probably the result of a more homogenous study population, and therefore may indicate an overfitting of their model. Based on the RMSE, Carreon et al al[22] concluded that individual patients' EQ-5D-3L utility values could not validly be predicted from their ODI scores. Although we agree with this conclusion, we would like to stress that a low RMSE does not necessarily mean that the models cannot be used in the context of a cost effectiveness analysis. This is true, when the bias surrounding the predicted utility values does not translate into relevant differences in incremental QALYs and the probability of the intervention being cost-effective compared to the control group (i.e., decision-based validity).[58] This may be explained by the fact that the bias is likely to be similar in the intervention and control groups, thereby not affecting incremental QALYs and CEACs.[59]

Strengths and limitations

To develop the models, a large sample of LBP patients from various settings (i.e., primary, secondary, and tertiary care) and with various complaint durations (i.e., subacute, and chronic LBP) was used, which increases both the reliability and generalisability of the models. Moreover, next to OLS models, Tobit models were used to account for the constrained range of utility values.[49,50] Although the added value of the Tobit model in this LBP population turned out to be rather limited, this might be different for LBP populations with milder symptoms, in which a larger share of patients is expected to report full health (i.e., a utility value of 1).

Our study also had some limitations. First, part of the sample was derived from two RCTs. Although RCT data may have limited generalisability, we chose to add these RCTs to our sample to create a more diverse sample and provide a better representation of the LPB population. Second, during the analysis, balanced data splitting was used to create the training and validation set. Although this balanced split provides better distribution of data then a random split, it might have been more appropriate to use K-fold cross validation.[60] Unfortunately, running the Tobit model using k-fold cross validation was not feasible as the R package for the Tobit model was not compatible with the K-fold package. In a post-hoc analysis we developed and validated the OLS models with k-fold cross validation, and this produced similar results as our main analysis (data not shown). We also expect this to be the case for the Tobit models. Third, EQ-5D-3L utilities were used instead of EQ-5D-5L utilities. This is a limitation because EQ-5D-5L is known to be more sensitive and therefore recommended in pharmacoeconomic guidelines. Nonetheless, some countries still use the EQ-5D-3L. Therefore, we preferred to use the current relatively large dataset with EQ-5D-3L utility values of nearly 20,000 patients for developing and validating the models, instead of using a relatively small

dataset with EQ-5D-5L. As the performance measures in the sensitivity analysis using the EQ-5D-5L reversed cross walk were comparable with those of the EQ-5D-3L version, we expect that EQ-5D-5L values can also be validly estimated using ODI scores. Fourth, the models were based on Dutch utility values. Previous research has shown that there are differences in utilities, QALYs, ICERs, and CEACs between countries due to the use of different value sets per country.[14] Therefore, we added the regression coefficients of models 2 and 5 for different countries in Appendix VI. These regression coefficients are based on the available value sets (tariffs) for different countries and can be used to calculate utility values and QALYs. Fifth, some data that were used were to assess the performance of the developed models in a trial-based cost effectiveness analysis setting were also part of the training set. However, as this was only a small percentage of the total training set (3.1%), we do not expect it to have influenced the validity of our finding that the difference between the estimated and true QALYs is small. Last, for assessing the performance of the developed models in a trial-based cost effectiveness analysis setting, we only used data of two clinical trials, both of which found the intervention far from being cost-effective. That is, the probability of the interventions being cost-effective was low regardless of the willingness to pay threshold. In datasets where the interventions' cost-effectiveness is less conclusive, even small differences in the probability of an intervention being cost-effective might impact the overall conclusion of a study. Further research in the form of a simulation study, using simulated data to examine the generalisability beyond the datasets, is needed to assess the performance of the developed models in a wide range of trial-based cost effectiveness analysis settings.

Implications for research and practice

Our findings suggest that predictive modelling can be used to estimate utility values from disease-specific measures, such as the ODI amongst LBP patients, when assessing incremental costs per QALY gained (as part of a cost effectiveness analysis) or differences in utilities between groups. This is helpful for assessing cost-effectiveness in trials that did not directly measure utilities. Given the relatively large RMSE (i.e., low absolute fit of the models) and the relatively low r-square value (i.e., low relative fit) it is strongly discouraged to use the developed models to estimate the utility values of individual patients. Further research is needed to validate the models in order to 1) assess whether these models yield comparable results in other empirical datasets on LBP interventions, especially in analysis on interventions that are expected not to be more conclusive in their cost-effectiveness, and 2) to improve their generalisability among different LBP patients by external validation in another sample. This study focussed on assessing the validity of predictive regression modelling in estimating EQ-5D-3L utility values from the ODI and the impact of these estimated utility values

on cost-effectiveness analysis. Results show that this is feasible for estimating QALYs and ICERs, but not for estimating individual utility scores. Further research is needed to explore whether other mapping methods, such as response mapping approaches like non-parametric and multinomial logistic regression,[16,17,55] result in better predictive accuracy in estimating individual utility values of preference-based measures, such as the EQ-5D. This is important because studies suggest these mapping methods might be better at preventing regression to the mean.[61] Additional research might not only result in more accurate estimated utility values but would also provide insight into the relative performance of different methods to estimate these values.

In the meantime, researchers can use the developed models in their cost effectiveness analysis when utility values are lacking. Of them, the OLS model (i.e., model 2) is recommended in samples in which only a small number of patients has a utility value of 1 at baseline or follow up measurement, whereas the Tobit model (i.e., model 5) is recommended in samples in which a substantial part of the sample has a utility score at baseline or at follow-up measurement. Although it seems possible to estimate utility values from disease-specific measures it is important to stress that it is still preferred to use preference-based quality of life measurements when setting up new studies.

Conclusion

Results of this study suggest that the ODI can be used to predict LBP patients' EQ-5D-3L utility values when the aim is to perform a cost effectiveness analysis for QALYs, if utility values are missing, meaning in order to compare difference between groups of patients. The models are not suitable for estimating utility values for individual patients. Further research is needed to validate the models in order to assess whether these models yield comparable results in other empirical datasets on LBP interventions, to improve generalisability of the estimated models, and to compare the performance of predictive modelling compared to a mapping approach for estimating utility values. In the meantime, researchers can use the developed models in their cost effectiveness analysis when utility values are lacking.

References

1. James SL, Abate D, Abate KH, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*. 2018;392(10159):1789-1858. doi:10.1016/S0140-6736(18)32279-7
2. MacNeela P, Doyle C, O'Gorman D, Ruane N, McGuire BE. Experiences of chronic low back pain: a meta-ethnography of qualitative research. *Health Psychol Rev*. 2015;9(1):63-82. doi:10.1080/17437199.2013.840951
3. Froud R, Patterson S, Eldridge S, et al. A systematic review and meta-synthesis of the impact of low back pain on people's lives. *BMC Musculoskelet Disord*. 2014;15(1):50. doi:10.1186/1471-2474-15-50
4. Ihlebæk C, Hansson TH, Lærum E, et al. Prevalence of low back pain and sickness absence: A ``borderline'' study in Norway and Sweden. *Scand J Public Health*. 2006;34(5):555-558. doi:10.1080/14034940600552051
5. Steenstra IA, Munhall C, Irvin E, 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;27(3):369-381. doi:10.1007/s10926-016-9666-x
6. Hush JM, Refshauge K, Sullivan G, De Souza L, Maher CG, McAuley JH. Recovery: What does this mean to patients with low back pain? *Arthritis Care Res (Hoboken)*. 2008;61(1):124-131. doi:10.1002/art.24162
7. Maniadakis N, Gray A. The economic burden of back pain in the UK. *Pain*. 2000;84(1):95-103. doi:10.1016/S0304-3959(99)00187-6
8. Lambeek LC, van Tulder MW, Swinkels ICS, Koppes LLJ, Anema JR, van Mechelen W. The Trend in Total Cost of Back Pain in the Netherlands in the Period 2002 to 2007. *Spine (Phila Pa 1976)*. 2011;36(13):1050-1058. doi:10.1097/BRS.0b013e3181e70488
9. Drummond M, Sculpher M, O'Brien B, Torrance G, Stoddart G. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd ed. Oxford University Press; 2005.
10. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. *Measuring and Valuing Health Benefits for Economic Evaluation*. Vol 1. Oxford University Press; 2016. doi:10.1093/med/9780198725923.001.0001
11. Froberg DG, Kane RL. Methodology for measuring health-state preferences–I: Measurement strategies. *J Clin Epidemiol*. 1989;42(4):345-354. doi:10.1016/0895-4356(89)90039-5
12. Hakkaart-van Roijen L, van der Linden N, Bouwmans C, Kanters T, Tan S. *Kostenhandleiding: Methodologie van Kostenonderzoek En Referentieprijzen Voor Economische Evaluaties in de Gezondheidszorg*.; 2015. Accessed August 30, 2021. https://www.zorginstituutnederland.nl/over-ons/publicaties/publicatie/2016/02/29/richtlijn-voor-het-uitvoeren-van-economische-evaluaties-in-de-gezondheidszorg
13. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. Published 2013. Accessed August 30, 2021. https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal- guidance/eq-5d-5l
14. van Dongen JM, Jornada Ben Â, Finch AP, et al. Assessing the Impact of EQ-5D Country-specific Value Sets on Cost-utility Outcomes. *Med Care*. 2021;59(1):82-90. doi:10.1097/MLR.0000000000001417
15. Gianola S, Frigerio P, Agostini M, et al. Completeness of Outcomes Description Reported in Low Back Pain Rehabilitation Interventions: A Survey of 185 Randomized Trials. *Physiotherapy Canada*. 2016;68(3):267-274. doi:10.3138/ptc.2015-30

16. Fayers PM, Hays RD. Should Linking Replace Regression When Mapping from Profile-Based Measures to Preference-Based Measures? *Value in Health*. 2014;17(2):261-265. doi:10.1016/j.jval.2013.1
17. Longworth L, Rowen D. Mapping to Obtain EQ-5D Utility Values for Use in NICE Health Technology Assessments. *Value in Health*. 2013;16(1):202-210. doi:10.1016/j.jval.2012.10.010
18. Chapman JR, Norvell DC, Hermsmeyer JT, et al. Evaluating Common Outcomes for Measuring Treatment Success for Chronic Low Back Pain. *Spine (Phila Pa 1976)*. 2011;36:S54-S68. doi:10.1097/BRS.0b013e31822ef74d
19. Fairbank J, Couper J, Davies J, O'Brien J. The Oswestry low back pain disability questionnaire. *Physiotherapy*. 1980;66(8):271-273.
20. Chiarotto A, Boers M, Deyo RA, et al. Core outcome measurement instruments for clinical trials in nonspecific low back pain. *Pain*. 2018;159(3):481-495. doi:10.1097/j.pain.0000000000001117
21. Clement RC, Welander A, Stowell C, et al. A proposed set of metrics for standardized outcome reporting in the management of low back pain. *Acta Orthop*. 2015;86(5):523-533. doi:10.3109/17453674.2015.1036696
22. Carreon LY, Bratcher KR, Das N, Nienhuis JB, Glassman SD. Estimating EQ-5D Values From the Oswestry Disability Index and Numeric Rating Scales for Back and Leg Pain. *Spine (Phila Pa 1976)*. 2014;39(8):678-682. doi:10.1097/BRS.0000000000000220
23. Mutubuki EN, Luitjens MA, Maas ET, et al. Predictive factors of high societal costs among chronic low back pain patients. *European Journal of Pain*. 2020;24(2):325-337. doi:10.1002/ejp.1488
24. Maas ET, Juch JN, Groeneweg JG, et al. Cost-effectiveness of minimal interventional procedures for chronic mechanical low back pain: design of four randomised controlled trials with an economic evaluation. *BMC Musculoskelet Disord*. 2012;13(1):260. doi:10.1186/1471-2474-13-260
25. Juch JNS, Maas ET, Ostelo RWJG, et al. Effect of Radiofrequency Denervation on Pain Intensity Among Patients With Chronic Low Back Pain. *JAMA*. 2017;318(1):68. doi:10.1001/jama.2017.7918
26. van Hooff ML, van Loon J, van Limbeek J, de Kleuver M. The Nijmegen Decision Tool for Chronic Low Back Pain. Development of a Clinical Decision Tool for Secondary or Tertiary Spine Care Specialists. *PLoS One*. 2014;9(8):e104226. doi:10.1371/journal.pone.0104226
27. van Dongen JM, van Hooff ML, Spruit M, de Kleuver M, Ostelo RWJG. Which patient-reported factors predict referral to spinal surgery? A cohort study among 4987 chronic low back pain patients. *European Spine Journal*. 2017;26(11):2782-2788. doi:10.1007/s00586-017-5201-9
28. van Hooff ML, van Dongen JM, Coupé VM, Spruit M, Ostelo RWJG, de Kleuver M. Can patient-reported profiles avoid unnecessary referral to a spine surgeon? An observational study to further develop the Nijmegen Decision Tool for Chronic Low Back Pain. *PLoS One*. 2018;13(9):e0203518. doi:10.1371/journal.pone.0203518
29. Apeldoorn AT, Ostelo RW, van Helvoirt H, Fritz JM, de Vet HC, van Tulder MW. The cost-effectiveness of a treatment-based classification system for low back pain: design of a randomised controlled trial and economic evaluation. *BMC Musculoskelet Disord*. 2010;11(1):58. doi:10.1186/1471-2474-11-58
30. Apeldoorn AT, Ostelo RW, van Helvoirt H, et al. A Randomized Controlled Trial on the Effectiveness of a Classification-Based System for Subacute and Chronic Low Back Pain. *Spine (Phila Pa 1976)*. 2012;37(16):1347-1356. doi:10.1097/BRS.0b013e31824d9f2b
31. Oosterhuis T, van Tulder M, Peul W, et al. Effectiveness and cost-effectiveness of rehabilitation after lumbar disc surgery (REALISE): design of a randomised controlled trial. *BMC Musculoskelet Disord*. 2013;14(1):124. doi:10.1186/1471-2474-14-124

32. Oosterhuis T, Ostelo RW, van Dongen JM, et al. Early rehabilitation after lumbar disc surgery is not effective or cost-effective compared to no referral: a randomised trial and economic evaluation. *J Physiother*. 2017;63(3):144-153. doi:10.1016/j.jphys.2017.05.016
33. EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy (New York)*. 1990;16(3):199-208. doi:10.1016/0168-8510(90)90421-9
34. Lamers L, Stalmeier P, McDonnell J, Krabbe P, van Busschbach J. 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.
35. Fairbank JCT, Pynsent PB. The Oswestry Disability Index. *Spine (Phila Pa 1976)*. 2000;25(22):2940-2953. doi:10.1097/00007632-200011150-00017
36. Hudson-Cook N, Tomes-Nicholson K, Breen A. Revised Oswestry disability questionnaire. In: Roland M, Jenner J, eds. *Back Pain: New Approaches to Rehabilitation and Education*. Manchester University Press; 1989:187-204.
37. Yeomans S, Liebenson C. Applying outcomes management to clinical practice. *Journal of the Neuromusculoskeletal System*. 1997;5(1):1-14.
38. Liebenson C. *Rehabilitation of the Spine – A Practitioner's Manual*. 2nd ed. Lippincott Williams & Wilkins; 2007.
39. van Hooff ML, Spruit M, Fairbank JCT, van Limbeek J, Jacobs WCH. The Oswestry Disability Index (Version 2.1a). *Spine (Phila Pa 1976)*. 2015;40(2):E83-E90. doi:10.1097/BRS.0000000000000683
40. Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA. Studies with pain rating scales. *Ann Rheum Dis*. 1978;37(4):378-381. doi:10.1136/ard.37.4.378
41. Boonstra AM, Stewart RE, Köke AJA, et al. Cut-Off Points for Mild, Moderate, and Severe Pain on the Numeric Rating Scale for Pain in Patients with Chronic Musculoskeletal Pain: Variability and Influence of Sex and Catastrophizing. *Front Psychol*. 2016;7. doi:10.3389/fpsyg.2016.01466
42. Husky MM, Ferdous Farin F, Compagnone P, Fermanian C, Kovess-Masfety V. Chronic back pain and its association with quality of life in a large French population survey. *Health Qual Life Outcomes*. 2018;16(1):195. doi:10.1186/s12955-018-1018-4
43. Horng YS, Hwang YH, Wu HC, et al. Predicting Health-Related Quality of Life in Patients With Low Back Pain. *Spine (Phila Pa 1976)*. 2005;30(5):551-555. doi:10.1097/01.brs.0000154623.20778.f0
44. Kovacs FM, Abraira V, Zamora J, Fernández C. The Transition from Acute to Subacute and Chronic Low Back Pain. *Spine (Phila Pa 1976)*. 2005;30(15):1786-1792. doi:10.1097/01.brs.0000172159.47152.dc
45. Lamé IE, Peters ML, Vlaeyen JWS, Kleef M V., Patijn J. Quality of life in chronic pain is more associated with beliefs about pain, than with pain intensity. *European Journal of Pain*. 2005;9(1):15-24. doi:10.1016/j.ejpain.2004.02.006
46. Bentsen SB, Wahl AK, Strand LI, Hanestad BR. Relationships between demographic, clinical and pain variables and health-related quality of life in patients with chronic low back pain treated with instrumented fusion. *Scand J Caring Sci*. 2007;21(1):134-143. doi:10.1111/j.1471-6712.2007.00440.x
47. Coste J, Lefrançois G, Guillemin F, Pouchot J. Prognosis and quality of life in patients with acute low back pain: Insights from a comprehensive inception cohort study. *Arthritis Care Res (Hoboken)*. 2004;51(2):168-176. doi:10.1002/art.20235
48. Hutcheson G, Sofroniou N. *The Multivariate Social Scientist*. 1st ed. SAGE Publications Ltd; 1999.
49. Tobin J. Estimation of relationships for limited dependent variables. *Econometrics*. 1958;26(1):24-36.
50. Austin PC, Escobar M, Kopec JA. The use of the Tobit model for analyzing measures of health status. *Quality of Life Research*. 2000;9(8):901-910. doi:10.1023/A:1008938326604
51. Smith G. Step away from stepwise. *J Big Data*. 2018;5(1):32. doi:10.1186/s40537-018-0143-6

52. Chowdhury MZI, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. *Fam Med Community Health*. 2020;8(1):e000262. doi:10.1136/fmch-2019-000262
53. Soer R, Reneman MF, Speijer BLGN, Coppes MH, Vroomen PCAJ. Clinimetric properties of the EuroQol-5D in patients with chronic low back pain. *The Spine Journal*. 2012;12(11):1035-1039. doi:10.1016/j.spinee.2012.10.030
54. Terluin B, van Marwijk HW, Adèr HJ, et al. The Four-Dimensional Symptom Questionnaire (4DSQ): a validation study of a multidimensional self-report questionnaire to assess distress, depression, anxiety and somatization. *BMC Psychiatry*. 2006;6(1):34. doi:10.1186/1471-244X-6-34
55. van Hout BA, Shaw JW. Mapping EQ-5D-3L to EQ-5D-5L. *Value in Health*. 2021;24(9):1285-1293. doi:10.1016/j.jval.2021.03.009
56. Euroqol. Cross-walk and reverse cross-walk. Accessed August 30, 2021. https://euroqol.org/support/tools/analysis-tools/cross-walk-reverse-cross-walk/
57. Euroqol. EQ-5D-5L version. Accessed August 30, 2021. https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/
58. Griffin S. *Dealing with Uncertainty in the Economic Evaluation of Health Care Technologies*. University of York; 2010.
59. Ben Â, Finch AP, Dongen JM, et al. Comparing the EQ-5D-5L crosswalks and value sets for England, the Netherlands and Spain: Exploring their impact on cost-utility results. *Health Econ*. 2020;29(5):640-651. doi:10.1002/hec.4008
60. Blum A, Kalai A, Langford J. Beating the hold-out. In: *Proceedings of the Twelfth Annual Conference on Computational Learning Theory*. ACM; 1999:203-208. doi:10.1145/307400.307439
61. Thompson NR, Lapin BR, Katzan IL. Mapping PROMIS Global Health Items to EuroQol (EQ-5D) Utility Scores Using Linear and Equipercentile Equating. *Pharmacoeconomics*. 2017;35(11):1167-1176. doi:10.1007/s40273-017-0541-1

Appendix I. Description studies included

MINT Study

The MINT study[23-25] assessed the effectiveness of radiofrequency denervation added to a standardized exercise program for patients with chronic low back pain. This study included patients with chronic LBP, receiving conservative treatment in a multidisciplinary pain clinic. This study was conducted at 16 multidisciplinary pain clinics in the Netherlands and had both a randomized and observational track. The randomized track consisted of three sub trails, namely the facet joint trial, the sacroiliac joint trial, and the combination trial (facet joint, sacroiliac joint, or the intervertebral disk). Patients were consecutively screened and were eligible when meeting the following criteria: a. pain considered to be related to the facet joint, sacroiliac joint, or a combination of the facet joint, sacroiliac joint, or intervertebral disk, aged 18 to 70 years, and no improvement in symptoms after conservative treatment. A total of 681 patients were included in the three randomized trails. Patients who were not willing to participate, or did not meet the inclusion criteria, were approached for the observational track of this study. In total 5168 patients were included in the observational track. Exclusion criteria for all trials were pregnancy, severe psychological problems, involvement in work-related conflicts or claims; body mass index higher than 35; or anticoagulant drug therapy or coagulopathy. Data was collection through surveys. For more details we refer to the original publications.

Nijmegen Decision Tool Study

In the Nijmegen Decision Tool Study (NDT study)[26,27] 47 indicators for a successful treatment outcome were assessed among chronic low back pain patients (CLBP), in order to compile a decision-support screening tool (NDT-CLBP) [28]. Patients were recruited at a Dutch orthopaedic hospital specialized in spine care, prior to their first consultation at the orthopaedic outpatient department. All consecutive low back pain patients were asked to complete the web-based questionnaire, which is part of routine practice. In total 14,859 patients with chronic LBP were included in this dataset. Patients were eligible when meeting the following criteria: experienced low back pain complaints for more than three months (i.e., CLBP) due to degenerative lumbar spine disorders (excluding trauma and tumour), had access to the internet, and were able to read and write Dutch. For more details we refer to the original publications.

Study of Apeldoorn et al.

The study of Apeldoorn et al.[29,20] assessed the cost-effectiveness of a modified version of Delitto's classification-based treatment approach compared with usual physical therapy care in patients with sub-acute and chronic LBP. This study included 156

patients with subacute and chronic LBP treated in a primary care setting. Patients were recruited by during their first contact with a physical therapist working in the region of Amsterdam. Patients were eligible when meeting the following criteria: LBP as the primary complaint (with or without associated leg pain), age between 18 and 65 years, current episode longer than 6 weeks, and able to read and write Dutch. Exclusion criteria were known- or suspected-specific LBP, severe radiculopathy, serious co-morbidity, and psychopathology. Data was collection through surveys. For more details we refer to the original publications.

REALISE Study

The REALISE study[31,32] concerned the assessment of effectiveness, and cost effectiveness of referral for early rehabilitation after lumbar disc surgery. This multicentre, randomised, controlled trial included 169 LPB patients with a herniated lumbar disc postoperatively treated in a primary care facility. Patients were referred to the research team by neurosurgeons and checked on eligibility by research nurses. Patients were eligible when meeting the following criteria: a herniated lumbar disc confirmed by magnetic resonance imaging (MRI) and signs of nerve root compression corresponding to the level of disc herniation, aged between 18 and 70 years, and were able to fill out questionnaires in Dutch themselves. Exclusion criteria were cauda equina syndrome, neurogenic claudication, co-morbidities of the lumbar spine, spinal surgery in the prior 12 months, contraindications to exercise therapy, pregnancy, or previous lumbar disc surgery at the same level and on the same side. Data was collection through surveys. For more details we refer to the original publications.

Table Appendix I. Baseline characteristics included studies (complete cases EQ-5D and ODI)

Age (years), mean (SD)
Sex, n (%)
Female
Male
Education level, n (%)
Low (no education, primary level education, lower vocational and lower secondary education)
Moderate (higher secondary education or undergraduate)
High (tertiary, university level, postgraduate)
Living together with a partner, n (%)
Yes
No
Type of low back pain, n (%)
Subacute (< 3 months)
Chronic (> 3 months)
Post-surgery, n (%)
Yes
No
Setting, n (%)
Primary care (i.e., physiotherapy clinics)
Secondary care (i.e., pain clinics)
Tertiary care (i.e., hospital)
NRS Pain, mean (SD)
Utility, mean (SD)
ODI, mean (SD)

SD = Standard Error, NRS = Numeric Rating Scale range 0-10, Utility range: -0.33 – 1, ODI: Oswestery Disability Index range: 0-100

Apeldoorn Study[29, 30] n=156	MINT Study [23-25] n=6,316	Nijmegen study [26-28] n=14,859	REALISE Study [31,32] n=169
42.5 (11.2)	56.2 (13.5)	53.5 (15.1)	47.3 (11.8)
89 (57.1)	3,576 (67.1)	8,695 (58.5)	98 (58.0)
67 (42.9)	1,757 (32.9)	6,164 (41.5)	71 (42.0)
27 (17.3)	1,892 (29.9)	3,922 (26.4)	37 (21.9)
61 (39.1)	2,406 (38.1)	6,967 (47.8)	97 (57.4)
68 (43.6)	823 (13.03)	3,403 (22.9)	35 (20.7)
119 (76.3)	4,663 (73.8)	11,118 (74.8)	125 (74.0)
37 (23.7)	1,593 (25.2)	3,741 (25.2)	44 (26.0)
32 (20.5)	3,601 (57.0)	423 (2.8)	0
124 (79.5)	1,682 (26.6)	14,436 (97.2)	169 (100.0)
0	0	0	169 (100.0)
156 (100.0)	6,316 (100.0)	14,859 (100.0)	0
156 (100.0)	0	0	169 (100.0)
0	6,316 (100.0)	0	0
0	0	14,859 (100.0)	
6.1 (1.8)	7.3 (1.6)	6.9 (2.0)	6.3 (2.6)
0.7 (0.2)	0.5 (0.3)	0.5 (0.3)	0.4 (0.3)
20.6 (13.0)	39.6 (14.6)	42.0 (15.4)	31.1 (14.3)

Appendix II. Regression coefficients model 1-6

Model 1. Ordinary Least Squares Regression with ODI total scores

Utility = 0.833 - 0.011*ODI total score + 0.002*age + 0.012*female + 0.015 *education middle + 0.021 *education high - 0.014 *no partner + 0.015* NRS moderate - 0.115 *NRS severe

		95% CI	
	Regression Coefficient (SE)	**2.5 %**	**97.5 %**
Intercept	0.833 (0.016)	0.807	0.857
ODI total score	-0.011 (0.000)	-0.011	-0.010
Age	0.002 (0.000)	0.001	0.002
Gender; female	0.012 (0.004)	0.003	0.019
Education; middle	0.015 (0.004)	0.006	0.024
Education; high	0.021 (0.006)	0.010	0.032
Partner; no partner	-0.014 (0.005)	-0.023	-0.005
NRS; moderate	0.015 (0.008)	-0.001	0.031
NRS; severe	-0.115 (0.009)	-0.131	-0.098

Model 2. Ordinary Least Squares Regression with ODI individual items scores (continuous)
Utility = 0.936 - 0.095*ODI1 - 0.044*ODI2 - 0.005*ODI3 - 0.019*ODI4 - 0.004*ODI5 - 0.008*ODI6 - 0.014*ODI7 - 0.033*ODI9 - 0.019*ODI10 + 0.002*age + 0.008*female + 0.019*education middle + 0.026*education high - 0.014*no partner - 0.066* secondary care - 0.051*tertiary care + 0.034*NRS moderate - 0.062*NRS severe

		95% CI	
	Regression Coefficient (SE)	**2.5 %**	**97.5 %**
Intercept	0.936 (0.024)	0.889	0.984
ODI1	-0.095 (0.003)	-0.099	-0.089
ODI2	-0.044 (0.002)	-0.049	-0.039
ODI3	-0.005 (0.002)	-0.009	-0.002
ODI4	-0.019 (0.002)	-0.023	-0.015
ODI5	-0.004 (0.002)	-0.008	-0.000
ODI6	-0.008 (0.002)	-0.012	-0.005
ODI7	-0.014 (0.002)	-0.018	-0.010
ODI9	-0.033 (0.002)	-0.039	-0.029
ODI10	-0.019 (0.002)	-0.023	-0.015
Age	0.002 (0.000)	0.001	0.002
Gender; female	0.008 (0.004)	0.000	0.015
Education; middle	0.019 (0.005)	0.009	0.027
Education; high	0.026 (0.005)	0.016	0.037
Partner; no partner	-0.014 (0.004)	-0.023	-0.006
Setting; secondary care	-0.066 (0.022)	-0.108	-0.024
Setting; tertiary care	-0.051 (0.021)	-0.093	-0.009
NRS; moderate	0.034 (0.008)	0.019	0.049
NRS; severe	-0.062 (0.008)	-0.079	-0.045

Model 3. Ordinary Least Squares Regression with ODI individual items scores (ordered)

Utility = 0.794 + 0.020*ODI1;1 - 0.004*ODI1;2 -0.138*ODI1;3 - 0.246*ODI1;4 - 0.247*ODI1;5 - 0.053 *ODI2;1 + 0.006*ODI2;2 - 0.106*ODI2;3 - 0.190*ODI2;4 - 0.146*ODI2;5 + 0.001*ODI3;1 - 0.001*ODI3;2 - 0.006*ODI3;3 - 0.012*ODI3;4 - 0.039*ODI3;5 - 0.017*ODI4;1 - 0.033*ODI4;2 -0.048*ODI4;3 - 0.069*ODI4;4 - 0.131*ODI4;5 + 0.004*ODI5;1 + 0.006*ODI5;2 - 0.009*ODI5;3 - 0.016*ODI5;4 - 0.026*ODI5;5 - 0.001*ODI6;1 - 0.005*ODI6;2 - 0.011*ODI6;3 - 0.024*ODI6;4 - 0.043*ODI6;5 - 0.003*ODI7;1 - 0.023*ODI7;2 - 0.036 *ODI7;3 - 0.049*ODI7;4 - 0.051*ODI7;5 - 0.024*ODI9;1 - 0.034*ODI9;2 - 0.093*ODI9;3 - 0.154*ODI9;4 - 0.153*ODI9;5 - 0.020*ODI10;1 - 0.042*ODI10;2 - 0.060*ODI10;3 - 0.079*ODI10;4 - 0.066*ODI10;5 + 0.002*age + 0.007*female + 0.017*education middle + 0.026*education high - 0.013*no partner - 0.088*secondary care - 0.078*tertiary care + 0.002*NRS moderate - 0.080*NRS severe

		95% CI	
	Regression Coefficient (SE)	**2.5 %**	**97.5 %**
Intercept	0.794 (0.028)	0.738	0.849
ODI1;1	0.020 (0.018)	-0.015	0.055
ODI1;2	-0.004 (0.017)	-0.037	0.029
ODI1;3	-0.138 (0.017)	-0.172	-0.104
ODI1;4	-0.246 (0.018)	-0.281	-0.211
ODI1;5	-0.247 (0.022)	-0.291	-0.204
ODI2;1	-0.053 (0.005)	-0.063	-0.043
ODI2;2	0.006 (0.006)	-0.109	-0.087
ODI2;3	-0.106 (0.008)	-0.122	-0.089
ODI2;4	-0.190 (0.015)	-0.221	-0.160
ODI2;5	-0.146 (0.040)	-0.224	-0.070
ODI3;1	0.001 (0.010)	-0.019	0.020
ODI3;2	-0.001 (0.010)	-0.021	0.020
ODI3;3	-0.006 (0.010)	-0.025	0.013
ODI3;4	-0.012 (0.010)	-0.032	0.008
ODI3;5	-0.039 (0.013)	-0.064	-0.013
ODI4;1	-0.017 (0.005)	-0.026	-0.007
ODI4;2	-0.033 (0.006)	-0.045	-0.021
ODI4;3	-0.048 (0.007)	-0.062	-0.033
ODI4;4	-0.070 (0.010)	-0.088	-0.050
ODI4;5	-0.131 (0.028)	-0.186	-0.077
ODI5;1	0.004 (0.008)	-0.012	0.020
ODI5;2	0.006 (0.008)	-0.009	0.021
ODI5;3	-0.009 (0.009)	-0.025	0.008
ODI5;4	-0.016 (0.010)	-0.036	0.005
ODI5;5	-0.026 (0.018)	-0.061	0.008
ODI6;1	-0.001 (0.011)	-0.022	0.021

Model 3. Continued

		95% CI	
	Regression Coefficient (SE)	**2.5 %**	**97.5 %**
ODI6;2	-0.005(0.011)	-0.027	0.016
ODI6;3	-0.011(0.011)	-0.032	0.011
ODI6;4	-0.024 (0.011)	-0.045	-0.003
ODI6;5	-0.043(0.013)	-0.070	-0.017
ODI7;1	-0.003(0.006)	-0.015	0.009
ODI7;2	-0.023(0.007)	-0.036	-0.010
ODI7;3	-0.036(0.008)	-0.108	-0.078
ODI7;4	-0.049(0.013)	-0.074	-0.025
ODI7;5	-0.051(0.016)	-0.082	-0.020
ODI9;1	-0.024(0.007)	-0.039	-0.009
ODI9;2	-0.034(0.007)	-0.048	-0.019
ODI9;3	-0.093(0.008)	-0.108	-0.078
ODI9;4	-0.154(0.011)	-0.175	-0.132
ODI9;5	-0.153(0.011)	-0.187	-0.119
ODI10;1	-0.020(0.008)	-0.037	-0.004
ODI10;2	-0.042(0.009)	-0.060	-0.024
ODI10;3	-0.060(0.010)	-0.079	-0.040
ODI10;4	-0.079 (0.011)	-0.101	-0.057
ODI10;5	-0.066(0.012)	-0.090	-0.041
Age	0.002(0.000)	0.001	0.002
Gender; female	0.007(0.004)	-0.001	0.014
Education; middle	0.017(0.004)	0.008	0.026
Education; high	0.026(0.005)	0.015	0.036
Partner; No partner	-0.013(0.004)	-0.021	-0.004
Setting; secondary care	-0.088 (0.022)	-0.131	-0.046
Setting; tertiary care	-0.078 (0.022)	-0.120	-0.035
NRS; moderate	0.002 (0.008)	-0.014	0.018
NRS; severe	-0.080 (0.009)	-0.096	-0.063

Model 4. Tobit with ODI total scores

Utility = 0.897 - 0.011*ODI total score + 0.002*age + 0.011*female + 0.015*education middle + 0.021*education high - 0.014*no partner - 0.058*secondary care - 0.058*tertiary care + 0.010*NRS moderate - 0.119*NRS severe

		95% CI	
	Regression Coefficient (SE)	**2.5 %**	**97.5 %**
Intercept	0.897 (0.025)	0.848	0.947
ODI total score	-0.011 (0.000)	-0.011	-0.011
Age	0.002 (0.000)	0.002	0.002
Gender; female	0.011 (0.004)	0.004	0.020
Education; middle	0.015 (0.004)	0.006	0.024
Education; high	0.021 (0.006)	0.010	0.032
Partner; no partner	-0.014 (0.005)	-0.023	-0.005
Setting; secondary care	-0.058 (0.023)	-0.104	-0.013
Setting; tertiary care	-0.058 (0.023)	-0.103	-0.013
NRS; moderate	0.010 (0.008)	-0.006	0.026
NRS; severe	-0.119 (0.009)	-0.136	-0.102

Model 5. Tobit with ODI individual items scores (continuous)

Utility = 0.961 - 0.096*ODI1 - 0.044*ODI2 - 0.005*ODI3 - 0.019*ODI4 - 0.005*ODI5 - 0.009*ODI6 - 0.014*ODI7 - 0.033*ODI9 - 0.019*ODI10 + 0.002*age + 0.008*female + 0.018*education middle + 0.026* education high - 0.014*no partner - 0.079*secondary care - 0.064* tertiary care + 0.029*NRS moderate - 0.066*NRS severe

		95% CI	
	Regression Coefficient (SE)	**2.5 %**	**97.5 %**
Intercept	0.961 (0.025)	0.913	1.009
ODI1	-0.096 (0.003)	-0.101	-0.091
ODI2	-0.044 (0.002)	-0.048	-0.040
ODI3	-0.005 (0.002)	-0.009	-0.002
ODI4	-0.019 (0.002)	-0.023	-0.015
ODI5	-0.005 (0.002)	-0.008	-0.000
ODI6	-0.009 (0.002)	-0.012	-0.005
ODI7	-0.014 (0.002)	-0.018	-0.010
ODI9	-0.033 (0.002)	-0.037	-0.029
ODI10	-0.019 (0.002)	-0.023	-0.015
Age	0.002 (0.000)	0.001	0.002
Gender; female	0.008 (0.004)	0.000	0.016
Education; middle	0.018 (0.005)	0.009	0.027
Education; high	0.026 (0.005)	0.016	0.037
Partner; no partner	-0.014 (0.004)	-0.023	-0.006
Setting; secondary care	-0.079 (0.022)	-0.123	-0.036
Setting; tertiary care	-0.064 (0.022)	-0.107	-0.021
NRS; moderate	0.029 (0.008)	0.013	0.045
NRS; severe	-0.066 (0.009)	-0.083	-0.049

Model 6. Tobit with ODI individual items scores (ordered)

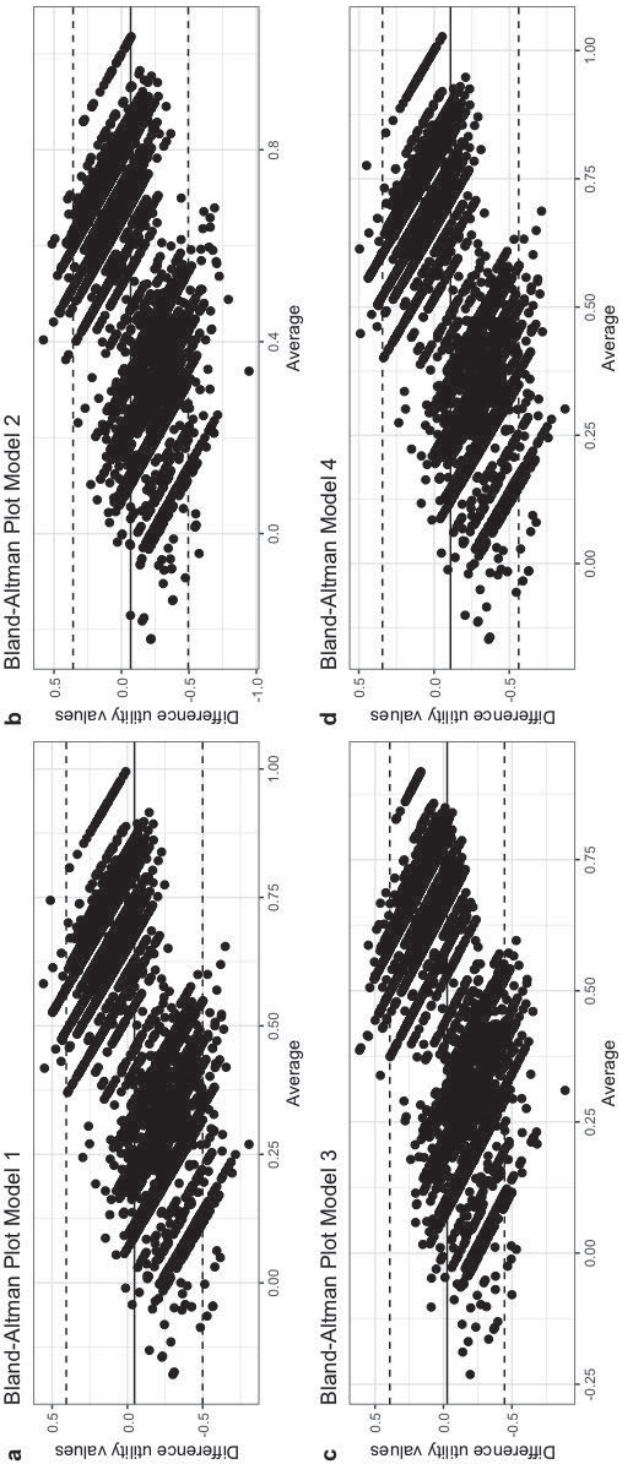
Utility = 0.831 - 0.006*ODI1;1 - 0.019*ODI1;2 - 0.153*ODI1;3 - 0.261*ODI1;4 - 0.262*ODI1;5 - 0.053*ODI2;1 - 0.098*ODI2;2 - 0.105*ODI2;3 - 0.190*ODI2;4 - 0.146*ODI2;5 - 0.002*ODI3;1 - 0.030*ODI3;2 - 0.009*ODI3;3 - 0.014*ODI3;4 - 0.041*ODI3;5 - 0.017*ODI4;1 - 0.033*ODI4;2 - 0.048*ODI4;3 - 0.069*ODI4;4 - 0.132*ODI4;5 - 0.004*ODI5;1 - 0.006*ODI5;2 - 0.009*ODI5;3 - 0.016*ODI5;4 -0.027*ODI5;5 - 0.002*ODI6;1 - 0.007*ODI6;2 - 0.013*ODI6;3 - 0.029*ODI6;4 - 0.045*ODI6;5- 0.003*ODI7;1 - 0.023*ODI7;2 - 0.036*ODI7;3 - 0.050*ODI7;4 - 0.051*ODI7;5 - 0.026*ODI9;1 - 0.036*ODI9;2 - 0.095*ODI9;3 - 0.155*ODI9;4 - 0.155*ODI9;5 -0.022*ODI10;1 - 0.043*ODi10;2 - 0.061*ODI10;3 - 0.080*ODI10;4 - 0.067*ODI10;5 + 0.0015134*age + 0.017 *education middle + 0.026* education high - 0.013*no partner - 0.099* secondary care - 0.089*tertiary care + 0.000 *NRS moderate - 0.081*NRS severe

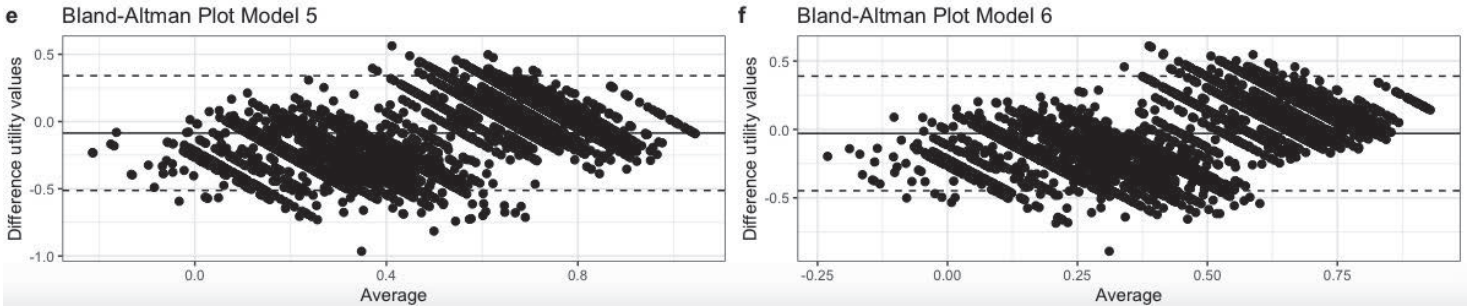
		95% CI	
	Regression Coefficient (SE)	**2.5 %**	**97.5 %**
Intercept	0.831 (0.029)	0.774	0.888
ODI1;1	0.006 (0.018)	-0.029	0.042
ODI1;2	-0.019 (0.017)	-0.054	0.015
ODI1;3	-0.153 (0.018)	-0.188	-0.120
ODI1;4	-0.261 (0.018)	-0.297	-0.226
ODI1;5	-0.262 (0.023)	-0.306	-0.218
ODI2;1	-0.053 (0.005)	-0.063	-0.043
ODI2;2	-0.098 (0.006)	-0.109	-0.087
ODI2;3	-0.105 (0.008)	-0.122	-0.089
ODI2;4	-0.190 (0.016)	-0.220	-0.159
ODI2;5	- 0.146 (-0.040)	-0.224	-0.069
ODI3;1	-0.002 (0.010)	-0.021	0.018
ODI3;2	-0.030 (0.010)	-0.024	0.018
ODI3;3	-0.009 (0.010)	-0.028	0.011
ODI3;4	-0.014 (0.010)	-0.034	0.006
ODI3;5	-0.041 (0.013)	-0.066	-0.016
ODI4;1	-0.017 (0.005)	-0.027	-0.007
ODI4;2	-0.033 (0.006)	-0.045	-0.021
ODI4;3	-0.048 (0.007)	-0.063	-0.037
ODI4;4	-0.069 (0.010)	-0.088	-0.051
ODI4;5	-0.132 (0.028)	-0.187	-0.078
ODI5;1	-0.004 (0.008)	-0.012	0.020
ODI5;2	-0.006 (0.008)	-0.010	0.021
ODI5;3	-0.009 (0.009)	-0.026	0.008
ODI5;4	-0.016 (0.011)	-0.037	0.004
ODI5;5	-0.027 (0.018)	-0.061	0.008
ODI6;1	-0.002 (0.011)	-0.023	0.020

Model 6. Continued

		95% CI	
	Regression Coefficient (SE)	**2.5 %**	**97.5 %**
ODI6;2	-0.007 (0.011)	-0.029	0.015
ODI6;3	-0.013 (0.011)	-0.034	0.090
ODI6;4	-0.029 (0.011)	-0.047	-0.004
ODI6;5	-0.045 (0.014)	-0.071	-0.019
ODI7;1	-0.003 (0.006)	-0.015	0.008
ODI7;2	-0.023 (0.007)	-0.036	-0.010
ODI7;3	-0.036 (-0.008)	-0.051	-0.021
ODI7;4	-0.050 (0.013)	-0.074	-0.025
ODI7;5	-0.051 (0.016)	-0.083	-0.020
ODI9;1	-0.026 (0.008)	-0.041	-0.011
ODI9;2	-0.036 (0.008)	-0.050	-0.021
ODI9;3	-0.095 (0.008)	-0.110	-0.080
ODI9;4	-0.155 (0.011)	-0.177	-0.133
ODI9;5	-0.155 (0.017)	-0.189	-0.121
ODI10;1	-0.022 (0.009)	-0.038	-0.005
ODI10;2	-0.043 (0.009)	-0.062	-0.025
ODI10;3	-0.061 (0.010)	-0.081	-0.041
ODI10;4	-0.080 (-0.011)	-0.103	-0.058
ODI10;5	-0.067 (0.012)	-0.091	-0.042
Age	0.002 (0.000)	0.001	0.002
Gender; female	-0.007 (0.004)	-0.001	0.015
Education; middle	0.017 (0.004)	0.008	0.025
Education; high	0.026 (0.005)	0.015	0.036
Partner; no partner	-0.013 (0.004)	-0.021	-0.004
Setting; secondary care	-0.099 (0.022)	-0.143	-0.056
Setting; tertiary care	-0.089 (0.022)	-0.132	-0.046
NRS; moderate	0.000 (0.008)	-0.016	0.016
NRS; severe	-0.081 (0.009)	-0.098	-0.064

Appendix III. Bland Altman plots estimates and actual utility values model 1-6 validation





Legend APPENDIX III

X-axis: Average measurement of the estimated and actual utility values, Y-axis: Difference in measurements between the two instruments.

Solid line: Average difference in measurements between the estimated and actual utility values, Dashed lines: 95% confidence interval limits for the average difference.

Appendix IV. Sensitivity analysis

SA 1 Mental Health
Model 1; OLS with ODI total scores
Model 2; OLS with ODI individual items scores (continuous)
Model 3; OLS with ODI individual items scores (ordered)
Model 4; Tobit with ODI total scores
Model 5; Tobit with ODI individual items scores (continuous)
Model 6; Tobit with ODI individual items scores (ordered)
OLS: Ordinary Least Squares Regression, ODI: Oswestery Disability Index, RMSE: root-mean-square error, R2: proportion of variation in the dependent variable, AIC: Akaike information criterion

SA 2 Living with partner
Model 1; OLS with ODI total scores
Model 2; OLS with ODI individual items scores (continuous)
Model 3; OLS with Stepwise Selection AIC with ODI sub scores (ordered)
Model 4; Tobit with ODI total scores
Model 5; Tobit with Stepwise Selection AIC with ODI sub scores (continuous)
Model 6; Tobit with Stepwise Selection AIC with ODI sub scores (ordered)
OLS: Ordinary Least Squares Regression, ODI: Oswestery Disability Index, RMSE: root-mean-square error, R2: proportion of variation in the dependent variable, AIC: Akaike information criterion

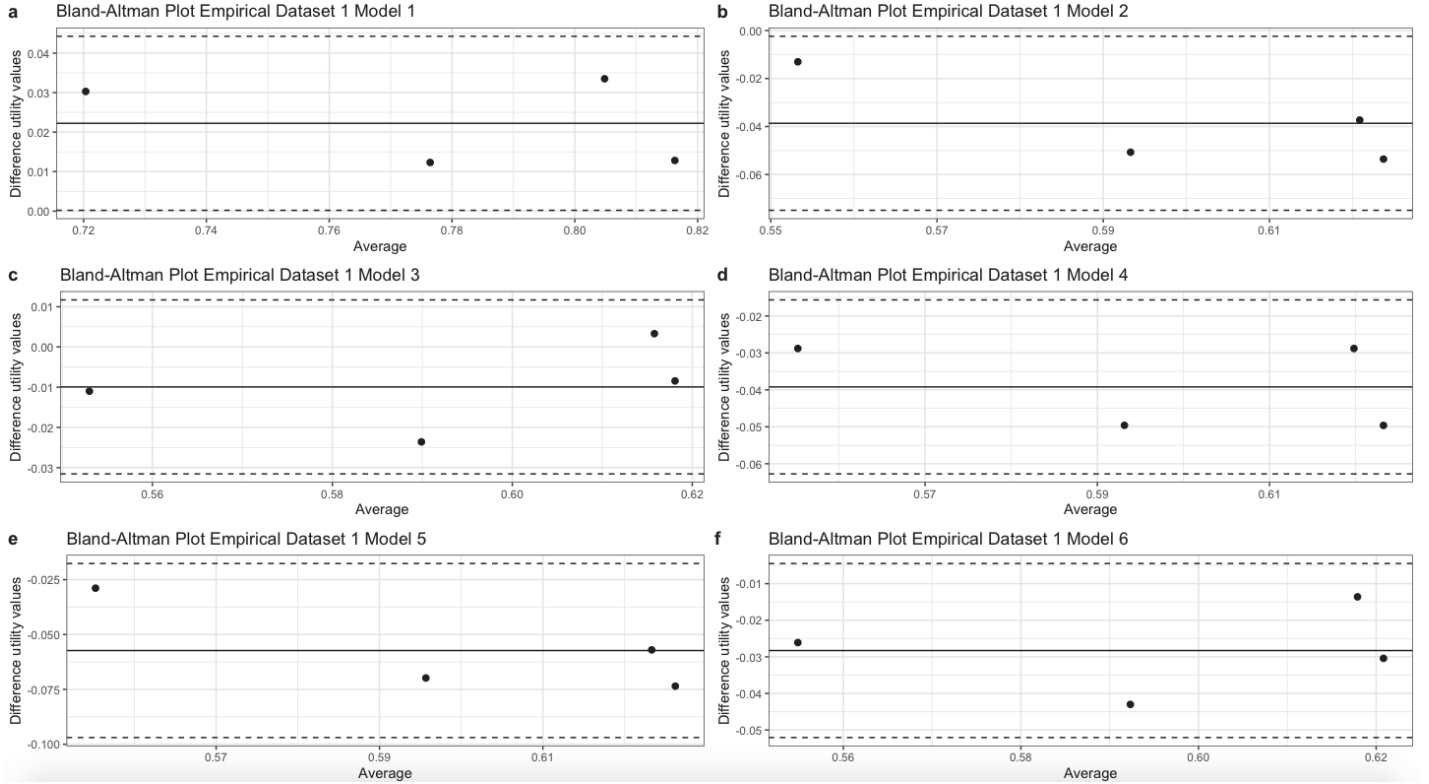
SA3 Cross walk EQ-5D-3
Model 1; OLS with ODI total scores
Model 2; OLS with ODI individual items scores (continuous)
Model 3; OLS with Stepwise Selection AIC with ODI sub scores (ordered)
Model 4; Tobit with ODI total scores
Model 5; Tobit with Stepwise Selection AIC with ODI sub scores (continuous)
Model 6; Tobit with Stepwise Selection AIC with ODI sub scores (ordered)
OLS: Ordinary Least Squares Regression, ODI: Oswestery Disability Index, RMSE: root-mean-square error, R2: proportion of variation in the dependent variable, AIC: Akaike information criterion

with mental health			without mental health		
RMSE	**R^2**	**AIC**	**RMSE**	**R^2**	**AIC**
0.21	0.46	-738.80	0.22	0.43	-617.10
0.20	0.49	-940.55	0.21	0.47	-802.74
0.20	0.52	-1004.90	0.20	0.49	-869.50
0.21	0.48	-699.13	0.22	0.44	-578.63
0.20	0.51	-904.42	0.21	0.48	-765.69
0.20	0.52	-960.76	0.20	0.49	-826.14

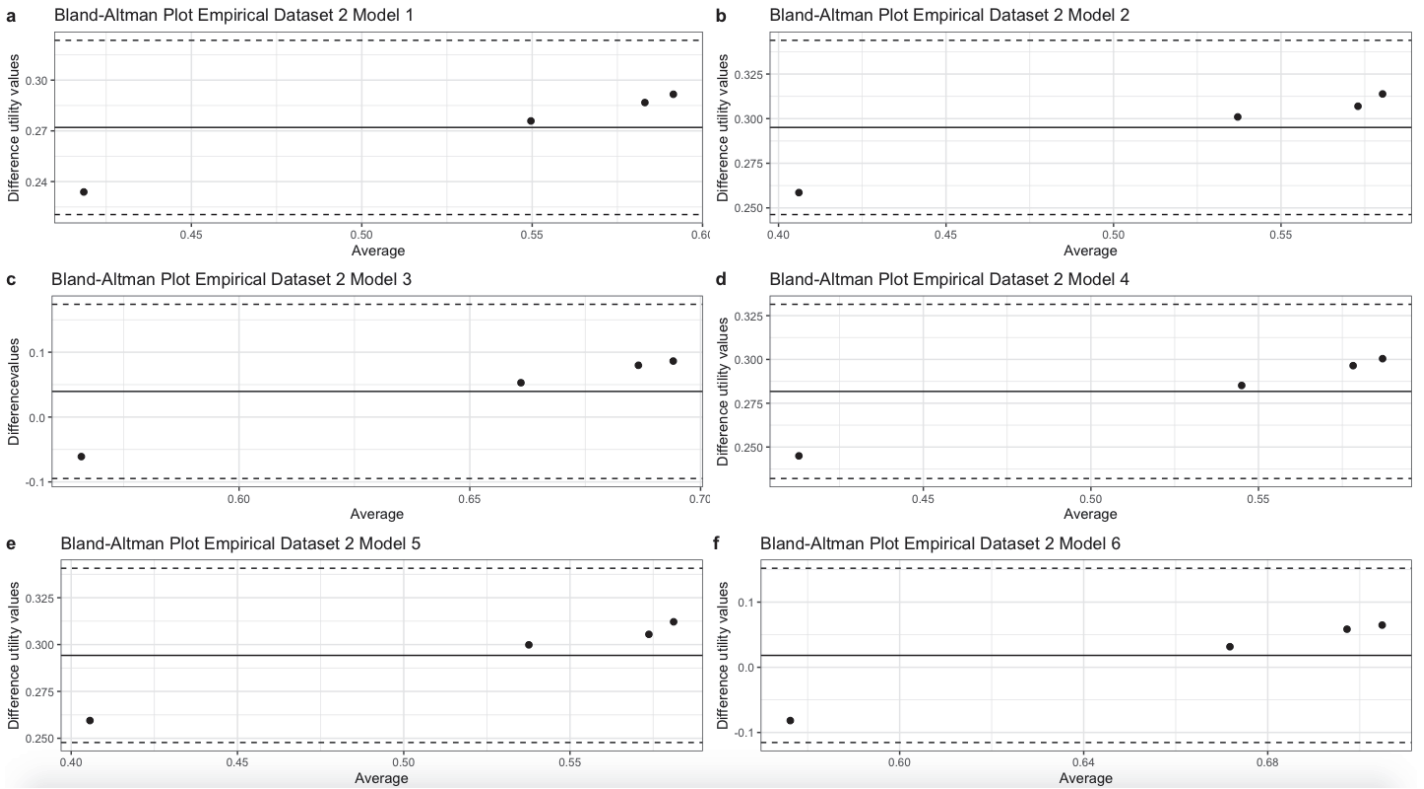
with variable partner			without variable partner		
RMSE	**R^2**	**AIC**	**RMSE**	**R^2**	**AIC**
0.22	0.45	-2326.48	0.22	0.45	-2318.62
0.21	0.50	-3423.24	0.21	0.50	-3401.74
0.21	0.51	-3768.53	0.21	0.51	-3762.27
0.22	0.46	-2054.46	0.22	0.46	-2061.91
0.21	0.50	-3155.95	0.21	0.50	-3164.37
0.21	0.51	-3467.56	0.21	0.51	-3473.60

EQ-5D-3L			EQ-5D-5L reversed cross walk		
RMSE	R^2	AIC	RMSE	R^2	AIC
0.22	0.45	-2326.48	0.15	0.49	-12150.58
0.21	0.50	-3423.24	0.15	0.53	-13158.25
0.21	0.51	-3769.51	0.14	0.54	-13412.33
0.22	0.45	-2061.91	0.15	0.49	-12156.93
0.21	0.50	-3164.37	0.15	0.53	-13158.25
0.21	0.51	-3474.88	0.14	0.54	-13412.33

Appendix V. Bland Altman plots estimated and actual utility values model 1-6 empirical datasets



X-axis: Average measurement of the estimated and actual utility values, Y-axis: Difference in measurements between the two instruments.



Solid line: Average difference in measurements between the estimated and actual utility values, Dashed lines: 95% confidence interval limits for the average difference.

Appendix VI. Regression coefficients per country for model 2 and 5

Regression coefficients Model 2	UK	Spain	Japan
Intercept	0.9216145	0.9629999	0.8061
ODI1	-0.1048651	-0.0912104	-0.03095
ODI2	-0.0527377	-0.0623234	-0.02753
ODI3	-0.0066237	-0.0083612	-0.003670
ODI4	-0.0261759	-0.0311805	-0.01816
ODI5	-0.0052813	-0.0059330	-0.004216
ODI6	-0.0114667	-0.0124294	-0.005125
ODI7	-0.0144788	-0.0125305	-0.005614
ODI9	-0.0338144	-0.0360919	-0.01856
ODI10	-0.0220272	-0.0247321	-0.01189
Age	0.0017108	0.0016504	0.0007201
Sex; female	0.0082351*	0.0127427	0.009772
Education; middle	0.0162531	0.0158689	0.007196
Education; high	0.0218355	0.0219954	0.01072
No partner	-0.0113913	-0.0085004*	-0.004576
Secondary care	-0.0820832	-0.0842040	-0.05519
Tertiary care	-0.0656978	-0.0748289	-0.05581
NRS; moderate	0.0388160	0.0359657	0.004944 *
NRS; severe	-0.0666840	-0.0567207	-0.01999
R^2 model	0.5155	0.5427	0.5339

* not significant

Zimbabwe	Germany	USA	South Korea	Denmark
0.8922	1.0250294	0.9124882	0.9344	0.8973506
-0.05299	-0.0975522	-0.0648293	-0.03990	-0.0757004
-0.03930	-0.0437269	-0.0386020	-0.03050	-0.0398935
-0.004353	-0.0052832	-0.0044403	-0.004936	-0.0045352
-0.01697	-0.0254695	-0.0176813	-0.02332	-0.0237435
-0.003785	-0.0032939	-0.0043486	-0.004155	-0.0056333
-0.006387	-0.0125239	-0.0070242	-0.006542	-0.0068254
-0.008621	-0.0119149	-0.0095046	-0.006776	-0.0132994
-0.02075	-0.0246979	-0.0230226	-0.02228	-0.0272122
-0.01395	-0.0167978	-0.0154409	-0.01668	-0.0190554
0.0009132	0.0012912	0.0011823	0.0008650	0.0013902
0.007933		0.0078843	0.01142	0.0095676
0.009566	0.0112980	0.0111634	0.009091	0.0129300
0.01294	0.0141464	0.0148199	0.01210	0.0153532
-0.005383	-0.0072071 *	-0.0088351	-0.003601	-0.0107631
-0.04963	-0.0788964	-0.0492312	-0.04828	-0.0481359
-0.04609	-0.0620024	-0.0411599	-0.04427	-0.0364969
0.01767	0.0395182	0.0239697	0.01842	0.0322145
-0.03398	-0.0612173	-0.0416691	-0.01741	-0.0391295
0.5631	0.4884	0.5289	0.5341	0.5173

Regression coefficients Model 2	France	Thailand	Canada
Intercept	0.8920794	0.8011386	0.8843
ODI1	-0.0698340	-0.0586726	-0.05749
ODI2	-0.0734759	-0.0487556	-0.03653
ODI3	-0.0081591	-0.0058457	-0.003727
ODI4	-0.0279455	-0.0252811	-0.01788
ODI5	-0.0059446	-0.0047299	-0.004710
ODI6	-0.0131530	-0.0103183	-0.005809
ODI7	-0.0119208	-0.0090852	-0.01049
ODI9	-0.0366947	-0.0248885	-0.02213
ODI10	-0.0176721	-0.0153155	-0.01320
Age	0.0012969	0.0009576	0.001065
Sex; female	0.0143939	0.0104631	0.007273
Education; middle	0.0188068	0.0105884	0.01101
Education; high	0.0325035	0.0162018	0.01475
No partner	-0.0100189	-0.0056038 *	-0.009769
Secondary care	-0.1155225	-0.0892417	-0.04946
Tertiary care	-0.1209327	-0.0879129	-0.04343
NRS; moderate	0.0080468 *	0.0131317	0.01803
NRS; severe	-0.0622104	-0.0425631	-0.03582
R^2 model	0.5757	0.5655	0.533

* not significant

China	Italy	Singapore	Taiwan	Argentina
0.8854	0.9449	0.8237519	0.7539217	0.9289
-0.04194	-0.05321	-0.0802764	-0.0730139	-0.05980
-0.04289	-0.03633	-0.0687301	-0.0657758	-0.02262
-0.005300	-0.004218	-0.0102091	-0.0081129	-0.004647
-0.01989	-0.02106	-0.0302459	-0.0267910	-0.02291
-0.003938	-0.005428	-0.0068003	-0.0055472	-0.004378
-0.007129	-0.004731	-0.0153680	-0.0110460	-0.01036
-0.007816	-0.009086	-0.0121587	-0.0120211	-0.007826
-0.02582	-0.02105	-0.0482907	-0.0404785	-0.01948
-0.01513	-0.01709	-0.0212139	-0.0224594	-0.01225
0.0008914	0.001171	0.0015746	0.0014902	0.0007053
0.01227	0.009246	0.0151897	0.0177048	0.006471
0.01318	0.007618	0.0255564	0.0215434	0.006118
0.02090	0.006850*	0.0457172	0.0349359	0.007121 *
-0.005817	-0.007207	-0.0126894	-0.0109666	
-0.06664	-0.02561 *	-0.1584970	-0.1102171	-0.09597
-0.06882	-0.02004*	-0.1554560	-0.1116358	-0.07828
0.008445	0.02737	-0.0012864*	0.0126286*	0.01409
0.03084	-0.02272	-0.0773060	-0.0584446	-0.03883
0.5753	0.515	0.557	0.5611	0.4742

Regression coefficients Model 2	Australia	Brazil	Chile
Intercept	0.9025891	0.7976	0.8235944
ODI1	-0.0801515	-0.02961	-0.0632935
ODI2	-0.0435985	-0.04156	-0.0591976
ODI3	-0.0046462	-0.005133	-0.0081320
ODI4	-0.0209143	-0.02042	-0.0271107
ODI5	-0.0040522	-0.004258	-0.0058750
ODI6	-0.0087226	-0.007792	-0.0099633
ODI7	-0.0118253	-0.005723	-0.0108634
ODI9	-0.0287889	-0.02157	-0.0375882
ODI10	-0.0169747	-0.01159	-0.0238120
Age	0.0013476	0.0006081	0.0013849
Sex; female	0.0072436	0.01137	0.0170097
Education; middle	0.0134860	0.009266	0.0180989
Education; high	0.0202586	0.01642	0.0289852
No partner	-0.0110786	-0.003376*	-0.0083945
Secondary care	-0.0705204	-0.08020	0.0893192
Tertiary care	-0.0601057	-0.08623	-0.0882755
NRS; moderate	0.0255881	-0.002747*	0.0146376
NRS; severe	-0.0541719	-0.02768	-0.0452285
R^2 model	0.53	0.5747	0.5681

* not significant

Hungary	Poland	Portugal	Sri Lanka	Sweden
0.9472674	0.9930241	0.7336	0.8370042	0.9092
-0.0601439	-0.0920958	-0.04206	-0.0599403	-0.02571
-0.0446507	-0.0406323	-0.05540	-0.0582604	-0.01638
-0.0053581	-0.0055945	-0.007265	-0.0084620	-0.003047
-0.0283114	-0.0253303	-0.02402	-0.0407823	-0.01112
-0.0073975	-0.0052011	-0.006597	-0.0059588	-0.002437
-0.0039895	-0.0089339	-0.009148	-0.0134058	-0.004826
-0.0110644	-0.0127106	-0.007490	-0.0106433	-0.004880
-0.0273724	-0.0271976	-0.02720	-0.0351063	-0.01726
-0.0257401	-0.0217994	-0.01842	-0.0248097	-0.007187
0.0015927	0.0014433	0.0009556	0.0012154	0.005130
0.0144952	0.0063456*	0.01633	0.0191124	0.005607
0.0098368	0.0104315	0.009909	0.0139406	0.009827
0.0064924*	0.0103657	0.01632	0.0208065	0.01747
-0.0085097	-0.0064591*	-0.08909		-0.005084
	-0.0467255	-0.09395	-0.0996952	-0.05385
	-0.0273513*	0.003993*	-0.0989189	-0.05025
0.0403118	0.0443564	0.005330	0.0213804	-0.0009108*
-0.0159125	-0.0466466	-0.03400	-0.0330848	-0.02.275
0.4754	0.5109	0.5689	0.552	0.5174

Regression coefficients Model 2	Trinidad Tobago	Belgium
Intercept	0.8883515	0.8447444
ODI1	-0.0390160	-0.0675222
ODI2	-0.0307738	-0.0396007
ODI3	-0.0035665	-0.0051525
ODI4	-0.0176425	-0.0182410
ODI5	-0.0046815	-0.0031076
ODI6	-0.0048159	-0.0086009
ODI7	-0.0071148	-0.0096681
ODI9	-0.0154423	-0.0282179
ODI10	-0.0122454	-0.0153905
Age	0.0008015	0.0011642
Sex; female	0.0076287	0.0087419
Education; middle	0.0048550	0.0159546
Education; high	0.0045773*	0.0248290
No partner	-0.0042406	-0.0097974
Secondary care	-0.0347116	-0.0818253
Tertiary care	-0.0323348	-0.0744276
NRS; moderate	0.0152583	0.0165661
NRS; severe	-0.0196063	-0.0503878
R^2 model	0.5291	0.5195

* not significant

Finland	Iran	Malaysia	New Zealand	Slovenia
0.7394	0.7589	0.9021951	0.8058	0.7929
-0.01883	-0.03214	-0.0403614	-0.05954	-0.03393
-0.03637	-0.06078	-0.0378891	-0.03396	-0.03881
-0.003874	-0.006695	-0.0047368	-0.004221	-0.004225
-0.01274	-0.01571	-0.0162989	-0.01650	-0.02606
-0.003825	-0.005023	-0.0029200	-0.002557*	-0.002369
-0.003055	-0.007477	-0.0069253	-0.007854	-0.01007
-0.005288	-0.007503	-0.0060276	-0.008642	-0.007424
-0.02356	-0.03182	-0.0224816	-0.02426	-0.02483
-0.01171	-0.01354	-0.0123636	-0.01277	-0.01078
0.0007940	0.0009362	0.0007950	0.0009665	0.0004864
0.01435	0.01686	0.0095461	0.007549	0.01211
0.01307	0.01732	0.0114116	0.01397	0.01465
0.02222	0.03269	0.0190117	0.02191	0.02579
-0.008083	-0.007903	-0.0053342	-0.008842	-0.006400
-0.05806	-0.09298	-0.0636079	-0.08000	-0.1092
-0.06886	-0.01078	-0.0650069	-0.07378	-0.1153
-0.007442	-0.01283	0.0077327 *	0.01226	-0.003784*
-0.02146	-0.04328	-0.0327250	-0.04561	-0.03126
0.471	0.5374	0.5658	0.5176	0.5395

Regression coefficients Model 5	UK	Spain	Japan
Intercept 1	0.9472972	0.9895059	0.8120
Intercept 2	-1.4601898	-1.5052735	-2.332
ODI1	-0.1060945	-0.0924513	-0.03121
ODI2	-0.0524924	-0.0620764	-0.02748
ODI3	-0.0068076	-0.0085525	-0.003711
ODI4	-0.0261440	-0.0311491	-0.01815
ODI5	-0.0056060	-0.0062683	-0.004289
ODI6	-0.0118429	-0.0128131	-0.005210
ODI7	-0.0144099	-0.0124585	-0.00560
ODI9	-0.0342032	-0.0364879	-0.01864
ODI10	-0.0218918	-0.0245948	-0.01186
Age	0.0016847	0.0016240	0.0007145
Sex; female	0.0085295	0.0130419	0.009837
Education; middle	0.0159807	0.0155870	0.007139
Education; high	0.0219722	0.0221218	0.01075
No partner	-0.0111703	-0.0082751*	-0.004524
Secondary care	-0.0960691	-0.0988243	-0.05854
Tertiary care	-0.0792959	-0.0890656	-0.05907
NRS; moderate	0.0336000	0.0306813	0.003827*
NRS; severe	-0.0708414	-0.0609302	-0.02088
R^2 model	0.5154651	0.5427006	0.533909

* not significant

Zimbabwe	Germany	USA	South Korea	Denmark
0.9032	1.0534441	0.9276111	0.9454	0.9154778
-2.084	-1.5363619	-1.9010432	-2.118	-1.7364138
-0.05350	-0.0988903	-0.0655513	-0.04043	-0.0765778
-0.03920	-0.0434683	-0.0384583	-0.03037	-0.0397189
-0.004431	-0.0054762	-0.0045482	-0.005017	-0.0046644
-0.01696	-0.0254331	-0.0176635	-0.02340	-0.0237235
-0.003924	-0.0036473*	-0.0045421	-0.004265	-0.0058678
-0.006547	-0.0129204	-0.0072481	-0.006690	-0.0070971
-0.008593	-0.0118338	-0.0094637	-0.006740	-0.0132512
-0.02092	-0.0251148	-0.0232536	-0.02251	-0.0274933
-0.01389	-0.0166507	-0.0153614	-0.01667	-0.0189594
0.0009022	0.0012619	0.0011671	0.0008651	0.0013719
0.008057		0.0080605	0.01135	0.0097820
0.009451	0.0110166	0.0110042	0.008913	0.0127378
0.01299	0.0142850	0.0149028	0.01213	0.0154533
-0.005286	-0.0069407*	-0.0087042		-0.0106022
-0.05567	-0.0944005	-0.0574566	-0.05516	-0.0579199
-0.05197	-0.0771188	-0.0491595	-0.05105	-0.0460041
0.01548	0.0338432	0.0208932	0.01602	-0.0460041
-0.03573	-0.0657410	-0.0441208	-0.01939	-0.0420809
0.5630955	0.4882995	0.5288496	0.533964	0.5172503

Regression coefficients Model 5	France	Thailand	Canada
Intercept 1	0.9105936	0.8118602	0.8965
Intercept 2	-1.6470017	-1.9013025	-1.997
ODI1	-0.0706407	-0.0591457	-0.05807
ODI2	-0.0733158	-0.0486618	-0.03641
ODI3	-0.0082881	-0.0059205	-0.003641
ODI4	-0.0279223	-0.0252677	-0.01786
ODI5	-0.0061698	-0.0048615	-0.004866
ODI6	-0.0134090	-0.0104702	-0.005989
ODI7	-0.0118732	-0.0090583	-0.01045
ODI9	-0.0369600	-0.0250451	-0.02231
ODI10	-0.0175793	-0.0152620	-0.01314
Age	0.0012792	0.0009472	0.001053
Sex; female	0.0145922	0.0105793	0.007414
Education; middle	0.0186230	0.0104814	0.01088
Education; high	0.0325879	0.0162499	0.01482
No partner	-0.0098760	-0.0055124*	-0.009664
Secondary care	-0.1261291	-0.0953430	-0.05617
Tertiary care	-0.1312972	-0.0938662	-0.04996
NRS; moderate	0.0045179*	0.0110833	0.01556
NRS; severe	-0.0650354	-0.0441987	-0.03779
R^2 model	0.5756387	0.565525	0.5329835

* not significant

China	Italy	Singapore	Taiwan	Argentina
0.8959	0.9608	0.8432469	0.7693004	0.9428
-2.098	-1.973	-1.5005430	-1.6125055	-1.924
-0.04240	-0.05398	-0.0811177	-0.0736964	-0.06042
-0.04280	-0.03618	-0.0685617	-0.0656403	-0.02250
-0.005374	-0.004334	-0.0103438	-0.0082196	-0.004745
-0.01988	-0.02105	-0.0302196	-0.0267717	-0.02288
-0.004068	-0.005639	-0.0070336	-0.0057366	-0.004546
-0.007278	-0.004974	-0.0156361	-0.0112651	-0.01055
-0.007789	-0.009042	-0.0121089	-0.0119813	-0.007793
-0.02597	-0.02130	-0.0485695	-0.0407062	-0.01968
-0.01508	-0.01700	-0.0211177	-0.0223827	-0.01218
0.0008813	0.001155	0.0015561	0.0014754	0.0006913
0.01238	0.009436	0.0153978	0.0178762	0.006627
0.01307	0.007445	0.0253681	0.0213923	0.005979*
0.02095	0.006931*	0.0458146	0.0350178	0.007173*
-0.005731	-0.007059	-0.0125426	-0.0108399	
-0.07265	-0.03417	-0.1697453	-0.1189331	-0.1040
-0.07468	-0.02836	-0.1664425	-0.1201396	-0.08606
0.006422*	0.02409	-0.0049764*	0.0096595*	0.01145
-0.03246	-0.02532	-0.0802601	-0.0608164	-0.04093
0.57526	0.51494	0.5570274	0.5610422	0.4741354

Regression coefficients Model 5	Australia	Brazil	Chile
Intercept 1	0.9205291	0.8039	0.8390576
Intercept 2	-1.7247032	-2.244	-1.7118986
ODI1	-0.0810006	-0.02987	-0.0639867
ODI2	-0.0434293	-0.04151	-0.0590597
ODI3	-0.0047727	-0.005176	-0.0082410
ODI4	-0.0208922	-0.02042	-0.0270919
ODI5	-0.0042776	-0.004334	-0.0060682
ODI6	-0.0089840	-0.007880	-0.0101857
ODI7	-0.0117777	-0.005708	-0.0108226
ODI9	-0.0290600	-0.02166	-0.0378188
ODI10	-0.0168807	-0.01156	-0.0237348
Age	0.0013295	0.0006022	0.0013700
Sex; female	0.0074489	0.01144	0.0171836
Education; middle	0.0132987	0.009206	0.0179436
Education; high	0.0203570	0.01645	0.0290658
No partner	-0.0109283	-0.003323*	-0.0082643
Secondary care	-0.0803519	-0.08387	-0.0980271
Tertiary care	-0.0696706	-0.08982	-0.0967684
NRS; moderate	0.0219669	-0.003909*	0.0116307*
NRS; severe	-0.0570602	-0.02861	-0.0476263
R^2 model	0.5300021	0.5746986	0.5680596

* not significant

Hungary	Poland	Portugal	Sri Lanka	Sweden
0.9594831	1.0188753	0.7424590	0.854156	0.9092
-1.6827221	-1.6065257	-1.9445409	-1.648065	-2.580
-0.0613315	-0.0933810	-0.0424434	-0.060690	-0.02571
-0.0443819	-0.0403758	-0.0553214	-0.058113	-0.01638
-0.0055609	-0.0057844	-0.0073265	-0.008583	-0.003047
-0.0283079	-0.0252997	-0.0240094	-0.040758	-0.01112
-0.0077569	-0.0055373	-0.0067070	-0.006172	-0.002347
-0.0043967	-0.0093227	-0.0092754	-0.013650	-0.004826
-0.0110356	-0.0126399	-0.0074682	-0.010600	-0.004880
-0.0277944	-0.0275985	-0.0273258	-0.035355	-0.01726
-0.0256004	-0.0216604	-0.0183716	-0.024724	-0.007187
0.0015644	0.0014163	0.0009470	0.001199	0.0005130
0.0147854	0.0066487*	0.0164295	0.019308	0.005607
0.0095554	0.0101440	0.0098232	0.013770	0.009827
0.0066940*	0.0104967	0.0163578		0.01742
-0.0082528	-0.0062182*		0.020878	-0.005084
	-0.0604633	-0.0941681	-0.109474	-0.05385
	-0.0406806*	-0.0989100	-0.108462	-0.05025
0.0353516	0.0389747	0.0023267*	0.018152	-0.0009108*
-0.0198679	-0.0509219	-0.0353305	-0.035651	-0.02275
0.475328	0.510803	0.5688545	0.5519598	0.5174331

Regression coefficients Model 5	Trinidad Tobago	Belgium
Intercept 1	0.8972	0.8585886
Intercept 2	-2.253	-1.8174829
ODI1	-0.03942	-0.0681537
ODI2	-0.03069	-0.0394748
ODI3	-0.003629	-0.0052490
ODI4	-0.01763	-0.0182228
ODI5	-0.004795	-0.0032777
ODI6	-0.004947	-0.0087993
ODI7	-0.007092	-0.009632
ODI9	-0.01558	-0.0284241
ODI10	-0.01220	-0.0153200
Age	0.0007929	0.0011506
Sex; female	0.007729	0.0088975
Education; middle	0.004763	0.0158167
Education; high	0.004618*	0.0249067
No partner	-0.004159*	-0.0096840
Secondary care	-0.03957	-0.0895791
Tertiary care	-0.03706	-0.0819805
NRS; moderate	0.01351	0.0138492
NRS; severe	-0.02100	-0.0525574
R^2 model	0.5290929	0.5194592

* not significant

Finland	Iran	Malaysia	New Zealand	Slovenia
0.7453	0.7684	0.9119	0.8161	0.8008
-2.232	-1.943	-2.191	-1.952	-2.091
-0.01908	-0.03253	-0.04079	-0.06001	-0.03424
-0.03632	-0.06070	-0.03780	-0.03387	-0.03875
-0.003914	-0.006760	-0.004805	-0.004292	-0.004277
-0.01273	-0.01570	-0.01629	-0.01649	-0.02604
-0.003899	-0.005138	-0.003039	-0.002683*	-0.002459
-0.003142	-0.007610	-0.007062	-0.008001	-0.01018
-0.005273	-0.007479	-0.006002	-0.008615	-0.007406
-0.02365	-0.03196	-0.02262	-0.02441	-0.02494
-0.01168	-0.01349	-0.01232	-0.01272	-0.01074
0.0007885	0.0009273	0.0007857	0.0009564	0.0004793
0.01442	0.01697	0.009653	0.007664	0.01219
0.01301	0.01723	0.01132	0.01387	0.01458
0.02225	0.03275	0.01906	0.02197	0.02583
-0.008030	-0.007824	-0.005256	-0.008757	-0.006341
-0.06151	-0.09850	-0.06916	-0.08582	-0.1138
-0.07223	-0.1132	-0.07043	-0.07945	-0.1198
-0.008571	-0.01461	0.005869*	0.01026*	-0.005173*
-0.02236	-0.04471	-0.03241	-0.04722	-0.03237
0.4709515	0.5374136	0.5657962	0.5176076	0.5394758


7

Mapping Oswestry Disability Index responses to EQ-5D utility values: are cost-utility results valid?

Value in Health. 2023. doi: 10.1016/j.jval.2023.01.020

Ben ÂJ, Pellekooren S, Bosmans JE, Ostelo RWJG, van Tulder MW, Maas ET, El Alili M, Huygen FJPM, Oosterhuis T, Apeldoorn AT, van Hooff ML, Dongen JM.

Abstract

Background

In the absence of the EQ-5D or another generic preference-based quality-of-life measure, utility values may be predicted using other measurement instruments, such as the frequently used Oswestry Disability Index (ODI) which is a condition-specific questionnaire on low back pain complaints. Evidence suggests that response mapping approaches perform better than regression models and might be better at preventing regression to the mean, because they aim to align the scales between instruments so that the distributions of their responses are matched. Hence, response mapping approaches might result in more accurate estimates of individual scores on the target instrument.

Objective

To develop and validate approaches for mapping ODI responses to EQ-5D-3L utility values and to evaluate the impact of using mapped utility values on cost-utility results compared to published regression models.

Methods

Three response mapping approaches were developed in a random sample of 70% of 18,692 patients with low back pain: non-parametric approach (Non-p), non-parametric approach excluding logical inconsistencies (Non-peLI), and ordinal logistic regression (OLR). Performance was assessed in the remaining 30% using R-square (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). To evaluate whether MAEs and their 95% limits of agreement (95%LA) were clinically relevant, a minimally clinically important difference (MCID) of 0.074 was used. Probabilities of cost-effectiveness estimated using observed and mapped utility values were compared in two economic evaluations.

Results

The Non-p performed best (R^2=0.43; RMSE=0.22; MAE=0.03; 95%LA=-0.40;0.47) compared to the Non-peLI (R^2=0.07; RMSE=0.29; MAE=-0.15; 95%LA=-0.63;0.34), and ORL (R^2=0.22; RMSE=0.26; MAE=0.02; 95%LA=-0.49;0.53). MAEs were lower than the MCID for the Non-p and OLR, but not for the Non-peLI. Differences in probabilities of cost-effectiveness ranged from 1-4% (Non-p), 0.1-9% (Non-peLI), and 0.1-20% (OLR).

Conclusions

Results suggest that the developed response mapping approaches are not valid for estimating individual patients' EQ-5D-3L utility values, and – depending on the approach - may considerably impact cost-utility results. The developed approaches did not perform better than previously published regression-based models and are therefore not recommended for use in economic evaluations.

Introduction

Low back pain (LBP) is a highly prevalent disabling condition that affects both physical and mental aspects of quality of life.[1,2] On top of that, LBP is associated with high healthcare and lost productivity costs.[3–5] Given the high disease burden of LBP, numerous healthcare interventions have been developed for preventing and/or treating LBP.[6–10] Before these interventions can be implemented on a large scale, information on their effectiveness and cost-effectiveness is needed to allocate healthcare resources as efficiently as possible.[11]

When evaluating the effectiveness of LBP interventions, the use of valid condition-specific, patient-reported outcome measures (PROMs) has been encouraged.[12,13] Two condition-specific PROMs are recommended for measuring physical functioning in studies conducted among LBP patients,[14] namely the Oswestry Disability Index (ODI)[15,16] and the Roland Morris Disability Questionnaire (RMDQ).[17] Of them, the ODI is the most widely used.[14]

When evaluating the cost-effectiveness of such interventions, health technology assessment agencies recommend using Quality-Life Adjusted-Years (QALYs).[18,19] QALYs are typically estimated using generic preference-based measures, such as the EQ-5D.[20] However, such questionnaires are not always included in clinical trials, as higher priority is generally given to condition-specific PROMs.[21] This issue is even more pronounced in real-world data (e.g., electronic health records), as these data are typically collected for clinical purposes only.[22] In these situations, mapping approaches might be used to estimate utility values based on condition-specific PROMs.[23–25]

Although EQ-5D and ODI seem to be conceptually linked,[26,27] it is unclear how strongly they are correlated. Estimates of their correlation vary from moderate[28] to high[28–31] for the total ODI score, and from low[28] to moderate for the ODI individual item scores.[28,29] Moreover, it is unclear whether the ODI is suitable for predicting missing EQ-5D utility values for LBP patients when EQ-5D scores are lacking.

Some studies have used ODI and RMDQ scores to predict EQ-5D utility values,[32–36] but none of them performed a qualitative assessment of their conceptual overlap. Moreover, all of these studies used regression modeling techniques to estimate utility values directly and did not include response mapping approaches.[37] In contrast to regression modeling techniques, response mapping approaches develop an algorithm to link item-responses between a source instrument (e.g., ODI) and a target instrument (e.g., EQ-5D). Based on the linked responses, the target instrument scores (e.g., utility scores) can be estimated.[23,37–40]

Evidence suggests that response mapping approaches perform better than regression models and might be better at preventing regression to the mean, because they align the scales between instruments so that the distributions of their responses are matched.[21,40,41] Hence, response mapping approaches might result in more accurate estimates of individual scores on the target instrument. This study therefore aimed to develop and assess the validity of response mapping approaches for mapping the ODI (source measure) to the EQ-5D (target measure) in LBP patients and to investigate the impact of using mapped utility values on cost-utility results. Then, the response mapping approaches' performance will be compared to that of published regression models[33] to assess their added value.

Methods

This study was performed in accordance with the MApping onto Preference-bases measures reporting Standards (MAPS) statement.[38] For the development and validation of the response mapping approaches, data used in this study were randomly split into a training set (70%) and validation set (30%) using the createDataPartition function of the caret R-package.[42]

Estimation Sample and Missing Data

Data used in this study were extracted from previously conducted studies among patients with sub-acute and chronic LBP in which both ODI and EQ-5D data were collected.[43–48] Complete baseline data from 18,692 out of 21,500 patients were used (Appendix I). Since the proportion of missing data on both instruments was <5%, exclusion of cases with missing observations was considered appropriate for handling missing data.[49]

Source and Target Measures

The *source measure* was the Dutch ODI version 2.1a.[16,50] It comprises ten items related to daily living (i.e., pain intensity, personal care, lifting, walking, sitting, standing, sleeping, sex life, social life, and traveling). Each item has six ordered response options, ranging from 0 (i.e., no limitations) to 5 (i.e., extreme limitations). The total ODI score is calculated by summing up the values of the response items, dividing it by the highest possible score over the collected ODI items (i.e., 50), and multiplying it by 100.[16] The total ODI score ranges from 0 to 100, with higher scores indicating higher levels of disability.[16] The ODI item on sex life is frequently excluded from applied studies, because participants might perceive it as inappropriate. Therefore, we excluded this item as well.[51,52] Please note that for the development of the response mapping approaches in this paper, only the response items were used and not the total ODI score.

The *target measure* was the EQ-5D-3L. The EQ-5D-3L includes five dimensions (i.e., mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) and three response levels (i.e., no problems, moderate problems, and severe problems),[20] resulting in a total of 243 health states (i.e., 3^5). To each health state, a utility value can be attached, which is preferably based on country-specific preferences. Utility values represent preferences for a certain health state on a scale anchored at 0 (i.e., death) and 1 (i.e., full health).[53] Negative values can also occur and indicate that a health state is valued as worse than death.[54] In this study, the Dutch EQ-5D-3L tariff was used to convert health states into utility values.[55]

Exploratory data analysis

An exploratory data analysis, including a qualitative and a quantitative assessment, was performed to evaluate the degree of conceptual overlap between the source and target measures. The qualitative assessment was based on two published studies[56,57] that evaluated the conceptual link between the ODI[56] and EQ-5D-3L,[57] and the International Classification of Functioning, Disability and Health (ICF), separately.[58] The quantitative assessment of the conceptual overlap was based on a Principal Component Analysis (PCA) using a varimax rotation.[59] The PCA was used to identify which ODI items were most strongly correlated with each of the five EQ-5D-3L dimensions, within a component and, thus, to reduce the number of ODI items from nine to five, which is required for the development of the response mapping approaches (i.e., one ODI item per EQ-5D-3L dimension). A correlation matrix, including Spearman correlations between ODI items and EQ-5D-3L dimensions was used as input for the PCA.[60] Subsequently, the factor loadings were estimated using a linear combination of the ODI items/EQ-5D-3L dimensions and the components. A varimax rotation was then applied, maximizing the sum of the variance of the squared loadings.[59]

Development of response mapping approaches

Three response mapping approaches were developed: 1) a non-parametric approach (Non-p), 2) a non-parametric approach excluding logical inconsistencies (Non-peLI), and 3) an ordinal logistic regression (ORL). The development of all approaches included two steps: 1) estimating transition probabilities between the ODI (source) and EQ-5D-3L (target) responses, and 2) estimating mapped utility values based on the transition probabilities. The three approaches differed in terms of the methodology used to estimate the transition probabilities, which is further described below.

Non-parametric approach

The Non-p was based on an adapted version of the non-parametric approach of van Hout et al.[61] First, transition probabilities were estimated between the ODI and EQ-5D-3L responses through cross-tabulation of the responses to the five selected ODI items

and their respective EQ-5D-3L dimensions (i.e., those that were found to measure the same construct). That information was used to calculate transition probabilities between responses of both instruments. Second, mapped utility values were estimated using the estimated transition probabilities using the Dutch tariff.[55]

Non-parametric approach excluding logical inconsistencies

Logical inconsistencies can occur when cross-tabulating responses of two instruments within their correlated constructs. For example, a respondent may indicate 0 on the ODI pain item (i.e., no limitation) and 3 on the pain/discomfort dimension of the EQ-5D-3L (i.e., severe pain). This may happen due to the different number of response levels between both instruments (i.e., ODI has six response levels, while EQ-5D-3L has three) or due to error. To check whether these logical inconsistencies impact the performance of the Non-p, observations were excluded from the dataset according to logical rules prior to estimating the transition probabilities. Logical inconsistencies were defined as any ODI response ≤3 simultaneously with any EQ-5D-3L response of 3, and any ODI response ≥4 simultaneously with any EQ-5D-3L response of 1. Then, mapped utility values were estimated using the estimated transition probabilities using the Dutch tariff.[55]

Ordinal Logistic Regression

The OLR is a response mapping approach in which transition probabilities from the ODI to the EQ-5D-3L were estimated indirectly by means of a logistic regression model, instead of cross-tabulation.[62] In this approach, five logistic models were constructed, each including one of the five EQ-5D-3L dimensions as a dependent variable and the ODI items that were found to measure the same construct as the five EQ-5D-3L dimensions as independent variables. Per logistic model, the log odds of the three EQ-5D-3L responses were modelled as a linear combination of the five ODI responses as shown in the formula below for the ODI item and EQ-5D-3L dimension:

$$ln\left(\frac{P(EQ5D3Lj=2}{P(EQ5D3Lj=1}\right) = b_{0j} + b_{i1}(ODI = 1) + b_{i2}(ODI = 2) + (\dots) + b_{i5}(ODI = 5)$$

Where *ln* is the natural logarithm, P is the probability of observing an EQ-5D-3L response level-1 compared to level-2 for dimension *j*. b_{0j} is the intercept and the coefficients b_{i1} to b_{i5} indicate the ODI response levels as dummy variables for ODI item *i*. The OLR model assumes that the dependent variable is ordered and that the probability of a specific response on the EQ-5D-3L is the same regardless of the ODI response level. For example, the probability of observing an EQ-5D-3L response level-1 compared to a response level-2 is the same for respondents who score 1 on the ODI item as for respondents who score 5 on the linked ODI item. To partially overcome this issue, the reference category

of the dependent variables was changed from level-1 to level-3 and level-2 to level-3. Additionally, during the model development we performed two sensitivity analyses: 1) a stratified analysis to investigate whether the transition probabilities differed by sex and age and 2) an ORL model including all ODI items as independent variables. If so, they would be included as covariates in the model. After obtaining the final transition probabilities, mapped utility values were estimated using the Dutch tariff.[55]

Validation

After developing the response mapping approaches on the training set, their performance was assessed in the validation set. Mapped utility values were compared with the observed values, valued using the EQ-5D-3L Dutch tariff.[63] The following performance measures were estimated: R-squared (R^2), root mean square error (RMSE), and mean absolute error (MAE). The R^2 assesses the explained variance of the response mapping approaches. The RMSE assesses their accuracy (i.e., a lower RMSE indicates better accuracy). The MAE represents the absolute difference between the mapped and observed EQ-5D-3L utility values. MAE and its 95%CI (further referred to as 95% limits of agreement, 95%LA) were additionally plotted using Bland-Altman plots.[64] To evaluate whether MAE and its respective 95%LA were clinically relevant, a minimally clinically important difference (MCID) of 0.074 for utility values was used as a threshold.[65]

Cost-utility analysis

Data from two RCTs[43,66] were used to assess the impact of the three response mapping approaches on cost-utility results (i.e., incremental QALY and probability of cost-effectiveness) compared to the "true" results (i.e., based on the observed utility values). Mapped and observed utility values were used to calculate QALYs. Subsequently, full economic evaluations were conducted using seemingly unrelated regressions to estimate incremental costs and QALYs. Incremental Cost-Effectiveness Ratios (ICERs) were calculated by dividing incremental costs by incremental QALYs. Uncertainty around incremental costs and QALYs was estimated using bootstrapping. Cost-Effectiveness Planes (CE-planes) were constructed, and the probability of cost-effectiveness was estimated for willingness-to pay thresholds of 0, 10.000, 30.000, and 50.000 euros/QALY.[67] Analyses were performed in R software, v4.1.2 and StataSE 16® (StataCorp LP, CollegeStation, TX, US).

Results

Estimation sample

The training set included 13,087 participants and the validation set 5,605 participants. The participants' characteristics in both datasets were similar (Table 1). The distribution of utility values in both sets followed a skewed and bimodal distribution (Figure 1).

Table 1. Characteristics of the participants assigned to the training and validation set

Characteristics	Complete data N=18,692	Training set n=13,087	Validation set n=5,605
Age (mean (SD))	53.9 (14.7)	53.9 (14.7)	54.0 (14.6)
Sex, female (n, %)	11,345 (60.7)	7,959 (60.8)	3,386 (60.4)
Level of education (n, %)			
Low	5,398 (28.9)	3,759 (28.7)	1,639 (29.2)
Intermediate	9,078 (48.6)	6,380 (48.8)	2,698 (48.1)
High	4,216 (22.6)	2,948 (22.5)	1,268 (22.6)
Living with a partner (n, %)	14,085 (75.4)	9,868 (75.4)	4,217 (75.2)
Type of LBP (n, %)			
Subacute (≤ 3 months)	3,248 (17.4)	2,245 (17.2)	1003 (17.9)
Chronic (> 3 months)	15,444 (82.6)	10,842 (82.8)	4,602 (82.1)
Post-surgery (n, %)	1,587 (8.5)	1,091 (8.3)	496 (8.8)
Setting (n, %)			
Primary care (i.e., physiotherapy clinics)	150 (0.8)	102 (0.8)	48 (0.9)
Secondary care (i.e., pain clinics)	4,123 (22.1)	2,876 (22.0)	1,247 (22.2)
Tertiary care (i.e., hospital)	14,419 (77.1)	10,109 (77.2)	4,310 (76.9)
NRS Pain (mean (SD))	6.99 (1.9)	7.0 (1.9)	7.0 (2.0)
Utility values (mean (SD))	0.467 (0.299)	0.466 (0.299)	0.469 (0.298)
ODI score[a] (mean (SD))	45.8 (17.1)	45.9 (17.1)	45.6 (17.1)

[a] excluding item ODI 8 sex life. LBP: Low Back Pain. NRS: Numeric Rating Scale (range 0-10). Utility values estimated by the EQ-5D-3L Dutch tariff (range -0.329 to 1). ODI: Oswestry Disability Scale (range 0-100). SD= Standard Deviation.

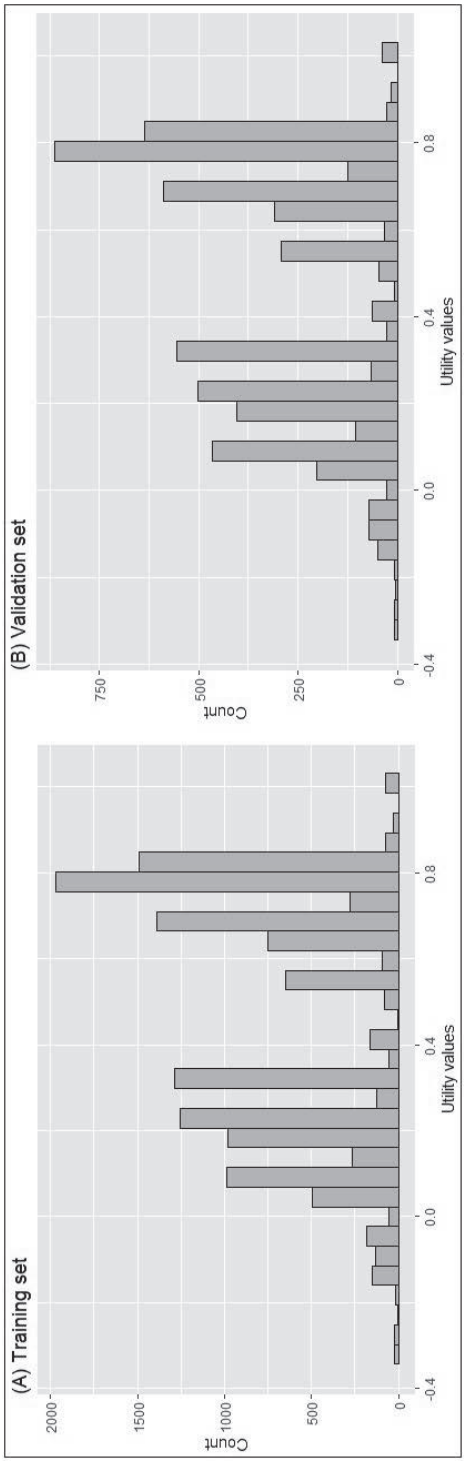


Figure 1. Distribution of utility values in the training and validation set

Exploratory analysis

The qualitative assessment of the conceptual overlap between both instruments showed that eight out of the fifteen ICF categories linked to the EQ-5D-3L (i.e., 53%) overlapped with those linked to the ODI (supplementary table 1). Overlap between both instruments was found for the ICF categories sensation of pain, mobility, walking, self-care, washing oneself, dressing, and recreation and leisure. No overlap was found for ICF categories related to Body Structures and Environmental Factors, as the EQ-5D-3L does not measure these specific concepts.

In the PCA rotated components, the five ODI items with the highest factor loadings into the EQ-5D-3L dimensions were: walking, sitting, personal care, pain intensity, and social life, respectively (Table 2). Despite the low factor loading of social life into the rotated component 5 (0.35), this ODI item was considered as a proxy for anxiety/depression.[68] These ODI items were used for the development of the response mapping approaches.

Table 2. ODI and EQ-5D-3L factor loadings into the five constructs of the Principal Component Analysis

	Components				
	RC 1	**RC 2**	**RC 3**	**RC 4**	**RC 5**
ODI Items					
Pain intensity	0.18	0.26	0.17	**0.80**	0.08
Personal care	0.31	0.19	**0.75**	0.22	0.11
Lifting	0.51	0.35	0.26	0.08	0.03
Walking	**0.80**	-0.01	0.17	0.14	0.07
Sitting	0.03	**0.82**	0.07	0.14	0.07
Standing	0.71	0.26	0.03	0.12	-0.04
Sleeping	0.09	0.57	0.18	0.34	0.03
Social life	0.52	0.40	0.20	0.12	**0.35**
Travelling	0.48	0.56	0.19	0.10	0.24
EQ-5D-3L dimensions					
Mobility	**0.67**	-0.11	0.17	0.19	0.12
Self-care	0.16	0.13	**0.88**	0.10	0.09
Usual Activities	0.39	**0.26**	0.19	0.12	0.40
Pain/discomfort	0.22	0.15	0.11	**0.82**	0.16
Anxiety/depression	0.04	0.06	0.08	0.13	**0.91**
Proportion of explained variance	*0.30*	*0.21*	*0.18*	*0.18*	*0.14*

ODI = Oswestry Disability Index. RC = Rotated component also referred to as constructs. Loadings represent the correlation coefficients between ODI items/ EQ-5D-3L dimensions and rotated components. Highlighted in bold are the ODI items with highest loadings into the same rotated component that of EQ-5D-3L.

Development of response mapping approaches

The cross-tabulation for the Non-p showed that ODI response levels were widely distributed within an EQ-5D-3L response level (Supplementary Table 2). For the development of the Non-peLI, a total of 8,159 patients were cross-tabulated after excluding observations with inconsistent responses (n=4,928) (Supplementary Table 3). As transition probabilities did not differ by sex, age, and also not when including for all ODI items (Supplementary Table 4), the final ORL model did not include those covariates.

Table 3 shows the transition probabilities for each response mapping approach. Transition probabilities were relatively similar between the Non-p and the Non-peLI but were considerably different between the non-parametric approaches and OLR. For example, the probability of observing response level-2 in the ODI pain intensity and in the EQ-5D-3L pain/discomfort was 0.8, 1.0, and 0, for the Non-p, Non-peLI, and OLR, respectively.

Validation

The mean observed utility value was 0.469 (SD=0.298, range -0.329 to 1) in the validation set. The mean mapped utility value obtained by the Non-p was 0.442 (SD=0.171, range=-0.057 to 0.807), by the Non-peLI 0.617 (SD=0.205, range=-0.106 to 0.824), and by the OLR 0.448 (SD=0.099, range=-0.010 to 0.545). The Non-p had a explained variance (R^2=0.43) and lower RMSE (0.22) compared to the Non-peLI and OLR (Table 4). OLR presented lower MAE (0.02) compared to the other response mapping approaches, however, it was not statistically different from those of Non-p and Non-peLI (Table 4).

The MAE was lower than the MCID for the Non-p and OLR, but not for the Non-peLI (Table 4). The Bland-Altman plots show that the MCID lies within the 95%LA for the three response mapping approaches (Figure 2).

Cost-utility analysis

For the Non-p, the difference in incremental QALY from the observed value was ≤0.01 in both economic evaluations. Differences in probabilities of cost-effectiveness between the Non-p and observed values ranged from 1-4% at different willingness-to-pay thresholds in both economic evaluations (Table 5).

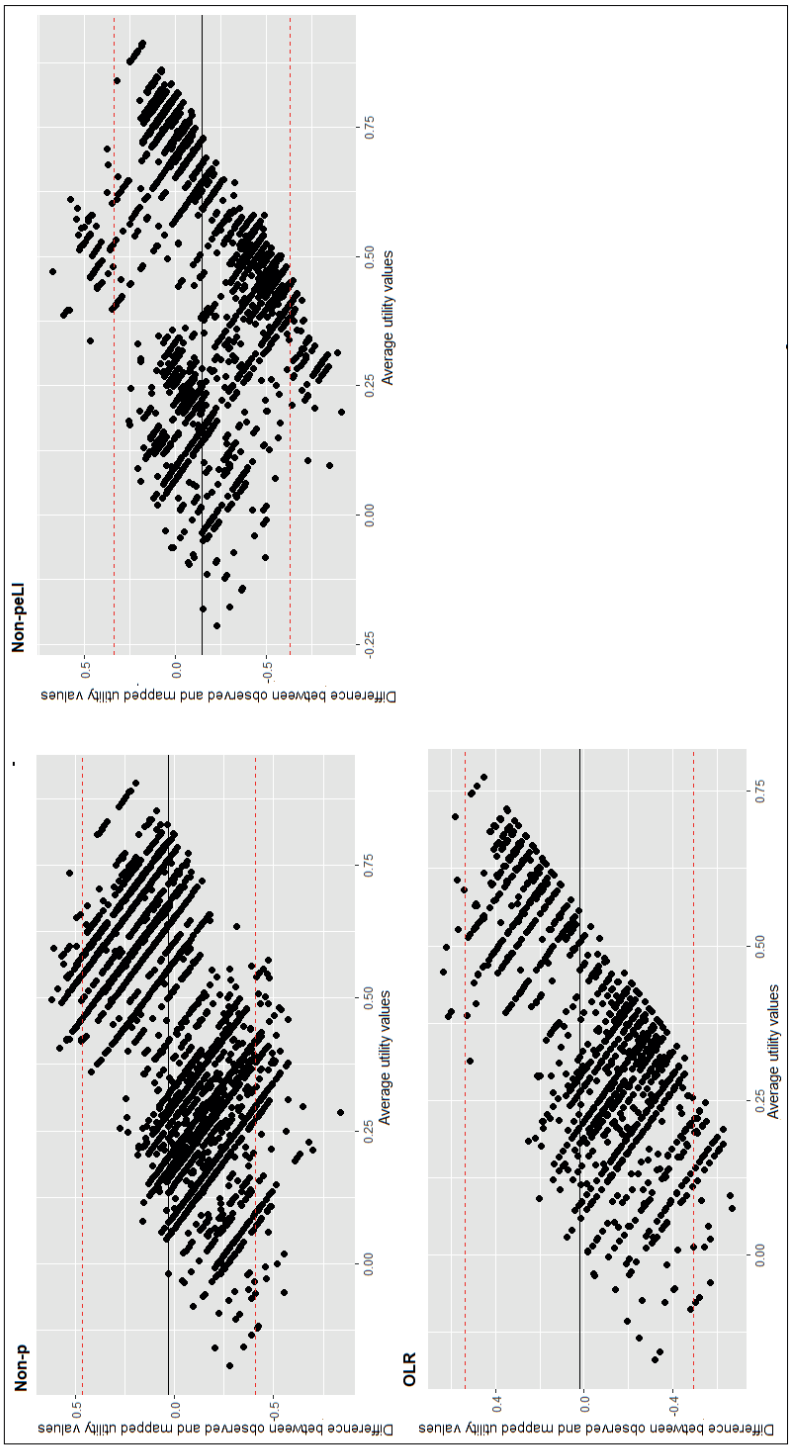


Figure 2. Bland-Altman plots for the three response mapping approaches

Table 3. Differences in transition probabilities for the three response mapping approaches

	Non-p (N=13,087)				
ODI→3L	**Mobility**	**Self-care**	**Usual Activities**	**Pain/ discomfort**	**Anxiety/ depression**
0→1	0.5	0.9	0.3	0.4	0.8
0→2	0.5	0.1	0.7	0.5	0.2
0→3	0	0	0	0.1	0
1→1	0.1	0.5	0.1	0.1	0.7
1→2	0.9	0.5	0.8	0.9	0.3
1→3	0	0	0.1	0	0
2→1	0.1	0.3	0.1	0	0.7
2→2	0.9	0.7	0.8	0.8	0.3
2→3	0	0	0.1	0.2	0
3→1	0	0.1	0	0	0.5
3→2	1	0.9	0.8	0.4	0.4
3→3	0	0	0.2	0.6	0.1
4→1	0	0	0	0	0.3
4→2	0.9	0.6	0.6	0.1	0.5
4→3	0.1	0.4	0.4	0.9	0.2
5→1	0	0	0	0	0.3
5→2	0.3	0.6	0.6	0	0.4
5→3	0.7	0.4	0.4	1	0.3

ODI = Oswestry Disability Index. Non-p: Non-parametric approach. Non-peLI: Non-parametric approach excluding logical inconsistencies. OLR: Ordinal Logistic Regression.

Table 4. Performance measures

	Performance training set (n=13,087)		
Response mapping approaches	**R^2**	**RMSE**	**MAE (95% CI or LA)**
Non-p	0.44	0.22	0.03 (-0.41; 0.46)
Non-peLI	0.06	0.29	-0.15 (-0.63; 0.33)
OLR	0.22	0.26	0.02 (-0.49; 0.53)

R^2: R square. RMSE: Root Mean Squared Error. MAE: mean absolute error. 95% CI: 95% confidence interval also referred to as 95% limits of agreement (LA).

Non-peLI (N=8,159)					OLR (N=13,087)				
Mobility	Self-care	Usual Activities	Pain/ discomfort	Anxiety/ depression	Mobility	Self-care	Usual Activities	Pain/ discomfort	Anxiety/ depression
0.6	0.9	0.3	0.4	0.9	0.8	0.6	0.8	0.5	0.6
0.4	0.1	0.7	0.6	0.1	0.2	0.4	0.1	0	0.4
0	0	0	0	0	0	0	0.1	0.5	0
0.2	0.6	0.2	0.1	0.8	0.8	0.6	0.8	0.5	0.6
0.8	0.4	0.8	0.9	0.2	0.2	0.4	0.1	0	0.4
0	0	0	0	0	0	0	0.1	0.5	0
0.1	0.3	0.1	0	0.7	0.2	0.4	0.1	0.5	0.4
0.9	0.7	0.9	1	0.3	0.8	0.6	0.8	0	0.6
0	0	0	0	0	0	0	0.1	0.5	0
0	0.1	0	0	0.6	0.2	0.4	0.1	0.5	0.4
1	0.9	1	1	0.4	0.8	0.6	0.8	0	0.6
0	0	0	0	0	0	0	0.1	0.5	0
0	0	0	0	0	0	0	0.1	0	0
1	0.6	0.7	0.1	0.7	0.2	0.6	0.1	0.5	0.6
0	0.4	0.3	0.9	0.3	0.8	0.4	0.8	0.5	0.4
0	0	0	0	0	0	0	0.1	0	0
0.3	0.5	0.6	0.1	0.5	0.2	0.6	0.1	0.5	0.6
0.7	0.5	0.4	0.9	0.5	0.8	0.4	0.8	0.5	0.4

ODI→3L represents a probability of a response in ODI is also observed in EQ-5D-3L. N = number of observations.

Performance validation set (n=5,605)		
R^2	RMSE	MAE (95% CI or LA)
0.43	0.22	0.03 (-0.40; 0.47)
0.07	0.29	-0.15 (-0.63; 0.34)
0.22	0.26	0.02 (-0.49; 0.53)

Non-p: Non-parametric approach. Non-peLI: Non-parametric approach excluding logical inconsistencies. OLR:Ordinal logistic regression.

Table 5. Cost-utility results

Models	DQALY (95% CI)	DC (95% CI)	ICER
RCT 1 n=86			
Observed values	-0.041 (-0.091; 0.009)	-110 (-1761; 1283)	2697
Non-p	-0.042 (-0.094; 0.008)	-110 (-1761; 1283)	2622
Non-peLI	-0.070 (-0.146; 0.007)	-110 (-1761; 1283)	1572
OLR	-0.019 (-0.066; 0.030)	-110 (-1761; 1283)	5697
RCT 2 n=424			
Observed values	-0.004 (-0.034; 0.027)	1576 (596; 2575)	-371566
Non-p	0.006 (-0.015; 0.027)	1576 (596; 2575)	280790
Non-peLI	0.0005 (-0.012; 0.013)	1576 (596; 2575)	3123224
OLR	-0.003 (-0.013; 0.005)	1576 (596; 2575)	-439092

RCT = Randomized Clinical Trial. N = number of observations in the analysis; DC= incremental costs; 95% CI = 95% confidence interval; DQALY= incremental QALY; ICER = Incremental Cost-Effectiveness Ratio; NE = northeast; SE = southeast; SW = southwest; NW = northwest;

For the Non-p and OLR, more pronounced differences in incremental QALY were found compared with the observed values (i.e, ≤0.03 and ≤0.02, respectively). As a consequence, larger differences in probabilities of cost-effectivenss between observed and estimated values were found for the Non-P and OLR compared to the Non-p (Table 5). That is, differences in probabilities of cost-effectiveness between the Non-peLI and observed values ranged from 0.1-9% while it ranged from 0.1-20% for the OLR.

Discussion

Main findings

This study developed and assessed the validity of three response mapping approaches to map the ODI to the EQ-5D-3L in LBP patients and investigated the impact of using mapped utility values on cost-utility results. The Non-p performed best (R^2=0.43; RMSE=0.22; MAE=0.03, 95%LA=-0.40 to 0.47) compared to the Non-peLI (R^2=0.07; RMSE=0.29; MAE=-0.15, 95%LA=-0.63 to 0.34), and the OLR (R^2=0.22; RMSE=0.26; MAE=0.02, 95%LA=-0.49 to 0.53). The relatively low explained variances of all response mapping approaches and the wide – and clinically relevant - limits of agreement between observed and mapped utility values may be explained by the identified suboptimal (53%) conceptual overlap between both instruments. This suggests that the developed response mapping approaches cannot be validly used for estimating individual patients'

Cost-effectiveness plane				Cost-effectiveness acceptability curve			
NE	SE	SW	NW	P_{CE} (0)	P_{CE} (10,000)	P_{CE} (30,000)	P_{CE} (50,000)
2%	4%	51%	42%	0.55	0.36	0.16	0.11
1%	5%	50%	44%	0.55	0.36	0.20	0.13
0%	3%	52%	45%	0.55	0.27	0.11	0.08
4%	18%	37%	41%	0.55	0.45	0.35	0.31
38%	0%	0%	62%	0.001	0.002	0.017	0.048
70%	0%	0%	30%	0.001	0.002	0.002	0.068
53%	0%	0%	47%	0.001	0.001	0.003	0.012
22%	0%	0%	78%	0.001	0.001	0.001	0.002

P_{CE} ( ) = probability that the intervention is cost-effective as compared to usual care with willingness-to-pay thresholds of 0, 10,000, 30,000, and 50,000 Euros. Non-p: Non-parametric approach. Non-peLI: Non-parametric approach excluding logical inconsistencies. OLR: Ordinal logistic regression.

utility values. The differences in probabilities of cost-effectiveness between observed and mapped utility values ranged from 1-4% for the Non-p, 0.1-9% for the Non-peLI, and 0.1-20% for the OLR. This indicates that use of mapped utility values may impact cost-utility results, particularly in the case of OLR.

Interpretation of the findings and comparison with the literature

Exploratory analysis

The identified suboptimal conceptual overlap between the ODI and EQ-5D-3L may be related to possible differences in the underlying construct of both instruments. Further research is needed to explore this issue, as systematic reviews on the content validity of various LBP-PROMs show that a clear definition of the constructs that the ODI and EQ-5D-3L intend to measure is missing.[69]

Response mapping approaches

For use in a cost-utility analysis, the Non-p appears to perform best compared to the other investigated response mapping approaches (i.e., Non-peLI, ORL). However, the performance of the Non-p was lower than that of a previously published OLS (R^2 =0.50, RMSE=0.21, MAE=0.05, 95%CI=-37;0.48) and a Tobit regression model (R^2 =0.50, RMSE=0.21, MAE=0.06, 95%CI=-36;0.48).[33] Both regression models were developed and validated using the same sample of LBP used in this study. The OLS model included EQ-5D-3L utility values as dependent variable and individual ODI items scores (assuming them to be continuous),

age, sex, education level (i.e., low, intermediate, and high), living with a partner (i.e.: yes or no), setting (i.e., primary, secondary, and tertiary care), and the numeric rating scale for pain intensity (categorized in low 0-3, moderate 4-6, and severe 7-10) as independent variables.[33] The Tobit model included the same set of variables as the OLS model, but accounted for possible ceiling effects in the dependent variable (i.e. utility values).[33]

The lower performance of the response mapping approach compared to that of regression models was unexpected, as the Non-p used the distribution of observed EQ-5D-3L responses for prediction instead of utility values,[21] and was, therefore, expected to be better at dealing with the EQ-5D-3L's well-known ceiling effect than an OLS. This finding may be explained by the fact that a relatively small proportion of the sample had a utility value of 1 (i.e., 111/18,692). The Non-p was also expected to prevent regression to the mean,[21,41,70] and thus to result in more accurately predicted utility values compared with the regression models. Again, this was not supported by our findings. A possible explanation for this might be that there were not many extreme scores, making regression to the mean less likely to occur.[71]

Despite the low performance of the response mapping approaches, the best performing approach (i.e., Non-p) resulted in relatively small differences in the probability of cost-effectiveness between observed and mapped values. These differences were similar to those of published regression models (1 to 5%).[33] This is probably due to the fact that bias is likely to be similar in both treatment groups, and hence only has a small on cost-utility outcomes.

In spite of previous evidence,[21,37,41,70] our results suggest that response mapping approaches are not necessarily preferred over regression models for mapping PROMs to EQ-5D-3L. That is, considering the better performance of the previously published regression models over the best performing response mapping approach (i.e., Non-p) in combination with the easier application of regression models, response mapping approaches do not seem to have added value compared to published regression models for the use in cost-utility analyses amongst LBP patients when EQ-5D-3L data were not collected.[33]

Strengths and limitations

One of the strengths of this study is the inclusion of a large, heterogeneous sample of LBP patients, with different underlying causes of LBP, different symptom durations (i.e., sub-acute, chronic), and different types of care (i.e., primary secondary, tertiary care). This improves the results' generalisability. Another strength is that the previously published regression models, we compared the developed mapping approaches to, were developed and validated in the same sample of LBP patients. This allowed for a

more valid comparison between regressions models and response mapping approaches for use in economic evaluations.

One of the main limitations was the use of the three-level version of the EQ-5D as target measure, whereas the EQ-5D-5L is now recommended. We used the EQ-5D-3L as target measure, since that allowed us to include a much larger sample compared to the available EQ-5D-5L datasets. Our use of the EQ-5D-3L, however, may have impacted the performance of the mapping approaches, as the relatively large difference in the number of response levels between the ODI (i.e. 6 levels) and EQ-5D-3L (i.e. 3 levels) made it challenging to properly link all responses.[31,72] Another limitation is that health states were collected from a Dutch sample and were valued using the Dutch tariff,[55] limiting generalizability of results to other countries.[73]

Implications for research and practice

Our findings suggests that use of the developed response mapping approaches is not recommended for predicting utility values from ODI to EQ-5D-3L in LBP patients.[33] That is, they had relatively low explained variances and a poor agreement between observed and mapped utility values. Moreover, they underperformed two previously published regression models in predicting EQ-5D-3L utility values from ODI scores, which are generally easier to implement than response mapping approaches.[33] Furthermore, even though previous studies suggest that response mapping approaches may be preferred over regression-based methods for mapping condition-specific instruments on the EQ-5D, our findings do not confirm this.[21,40,41] Further research is required to assess whether this is related to the relatively low conceptual overlap between the ODI and EQ-5D-3L and/or whether the preferred method might differ across diseases and/or PROMs. Investigation of the performance of response mapping approaches in other countries/populations is needed, especially in patient populations with better health states, and using the more sensitive version of the EQ-5D, i.e., the EQ-5D-5L.

Conclusion

Results suggest that the developed response mapping approaches are not valid for estimating individual patients' EQ-5D-3L utility values, and – depending on the approach - may considerably impact cost-utility results. The developed approaches did not perform better than previously published regression-based models and are therefore not recommended for use in economic evaluations. Further research comparing response mapping approaches and regression-based models for estimating EQ-5D utility values is needed.

References

1. Hartvigsen J, Hancock MJ, Kongsted A, et al. What low back pain is and why we need to pay attention. *The Lancet*. 2018;391(10137):2356-2367. doi:10.1016/S0140-6736(18)30480-X
2. James SL, Abate D, Abate KH, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*. 2018;392(10159):1789-1858. doi:10.1016/S0140-6736(18)32279-7
3. Dutmer AL, Schiphorst Preuper HR, Soer R, et al. Personal and Societal Impact of Low Back Pain. *Spine (Phila Pa 1976)*. 2019;44(24):E1443-E1451. doi:10.1097/BRS.0000000000003174
4. Kim LH, Vail D, Azad TD, et al. Expenditures and Health Care Utilization Among Adults With Newly Diagnosed Low Back and Lower Extremity Pain. *JAMA Netw Open*. 2019;2(5):e193676. doi:10.1001/jamanetworkopen.2019.3676
5. Carregaro RL, da Silva EN, van Tulder M. Direct healthcare costs of spinal disorders in Brazil. *Int J Public Health*. 2019;64(6):965-974. doi:10.1007/s00038-019-01211-6
6. Furlan AD, Giraldo M, Baskwill A, Irvin E, Imamura M. Massage for low-back pain. *Cochrane Database of Systematic Reviews*. 2015;2015(9). doi:10.1002/14651858.CD001929.pub3
7. Yamato TP, Maher CG, Saragiotto BT, et al. Pilates for low back pain. *Cochrane Database of Systematic Reviews*. 2015;2015(7). doi:10.1002/14651858.CD010265.pub2
8. Ebadi S, Henschke N, Forogh B, et al. Therapeutic ultrasound for chronic low back pain. *Cochrane Database of Systematic Reviews*. 2020;2020(7). doi:10.1002/14651858.CD009169.pub3
9. Maas ET, Ostelo RW, Niemisto L, et al. Radiofrequency denervation for chronic low back pain. *Cochrane Database of Systematic Reviews*. 2015;2015(10). doi:10.1002/14651858.CD008572.pub2
10. Choi BK, Verbeek JH, Tam WWS, Jiang JY. Exercises for prevention of recurrences of low-back pain. *Cochrane Database of Systematic Reviews*. Published online January 20, 2010. doi:10.1002/14651858.CD006555.pub2
11. van Dongen JM, Ketheswaran J, Tordrup D, Ostelo RWJG, Bertollini R, van Tulder MW. Health economic evidence gaps and methodological constraints in low back pain and neck pain: Results of the Research Agenda for Health Economic Evaluation (RAHEE) project. *Best Pract Res Clin Rheumatol*. 2016;30(6):981-993. doi:10.1016/j.berh.2017.09.001
12. Ostelo RWJG, de Vet HCW. Clinically important outcomes in low back pain. *Best Pract Res Clin Rheumatol*. 2005;19(4):593-607. doi:10.1016/j.berh.2005.03.003
13. Chiarotto A, Ostelo RW. Clinimetrics: A core outcome measurement set for low back pain. *J Physiother*. 2020;66(1):58. doi:10.1016/j.jphys.2019.07.011
14. Chiarotto A, Boers M, Deyo RA, et al. Core outcome measurement instruments for clinical trials in nonspecific low back pain. *Pain*. 2018;159(3):481-495. doi:10.1097/j.pain.0000000000001117
15. Froud R, Patterson S, Eldridge S, et al. A systematic review and meta-synthesis of the impact of low back pain on people's lives. *BMC Musculoskelet Disord*. 2014;15(1):50. doi:10.1186/1471-2474-15-50
16. Fairbank JCT, Pynsent PB. The Oswestry Disability Index. *Spine (Phila Pa 1976)*. 2000;25(22):2940-2953. doi:10.1097/00007632-200011150-00017
17. Roland M, Morris R. A Study of the Natural History of Back Pain. *Spine (Phila Pa 1976)*. 1983;8(2):141-144. doi:10.1097/00007632-198303000-00004
18. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. Published 2013. Accessed August 30, 2021. https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal- guidance/eq-5d-5l

19. Zorginstituut Nederland. *Richtlijn Voor Het Uitvoeren van Economische Evaluaties in de Gezondheidszorg*.; 2016. Accessed March 28, 2018. https://www.zorginstituutnederland.nl/publicaties/publicatie/2016/02/29/richtlijn-voor-het-uitvoeren-van-economische-evaluaties-in-de-gezondheidszorg
20. EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy (New York)*. 1990;16(3):199-208. doi:10.1016/0168-8510(90)90421-9
21. Fayers PM, Hays RD. Should Linking Replace Regression When Mapping from Profile-Based Measures to Preference-Based Measures? *Value in Health*. 2014;17(2):261-265. doi:10.1016/j.jval.2013.12.002
22. Fatoye F, Gebrye T, Odeyemi I. Real-world incidence and prevalence of low back pain using routinely collected data. *Rheumatol Int*. 2019;39(4):619-626. doi:10.1007/s00296-019-04273-0
23. Longworth L, Rowen D. Mapping to Obtain EQ-5D Utility Values for Use in NICE Health Technology Assessments. *Value in Health*. 2013;16(1):202-210. doi:10.1016/j.jval.2012.10.010
24. Mukuria C, Rowen D, Harnan S, et al. An Updated Systematic Review of Studies Mapping (or Cross-Walking) Measures of Health-Related Quality of Life to Generic Preference-Based Measures to Generate Utility Values. *Appl Health Econ Health Policy*. 2019;17(3):295-313. doi:10.1007/s40258-019-00467-6
25. Hernández Alava M, Wailoo A, Pudney S, Gray L, Manca A. Mapping clinical outcomes to generic preference-based outcome measures: development and comparison of methods. *Health Technol Assess (Rockv)*. 2020;24(34):1-68. doi:10.3310/hta24340
26. Solberg TK, Olsen JA, Ingebrigtsen T, Hofoss D, Nygaard ØP. Health-related quality of life assessment by the EuroQol-5D can provide cost-utility data in the field of low-back surgery. *European Spine Journal*. 2005;14(10):1000-1007. doi:10.1007/s00586-005-0898-2
27. Johnsen LG, Hellum C, Nygaard ØP, et al. Comparison of the SF6D, the EQ5D, and the oswestry disability index in patients with chronic low back pain and degenerative disc disease. *BMC Musculoskelet Disord*. 2013;14(1):148. doi:10.1186/1471-2474-14-148
28. Mueller B, Carreon LY, Glassman SD. Comparison of the EuroQOL-5D With the Oswestry Disability Index, Back and Leg Pain Scores in Patients With Degenerative Lumbar Spine Pathology. *Spine (Phila Pa 1976)*. 2013;38(9):757-761. doi:10.1097/BRS.0b013e31827ab803
29. Carreon LY, Bratcher KR, Das N, Nienhuis JB, Glassman SD. Estimating EQ-5D Values From the Oswestry Disability Index and Numeric Rating Scales for Back and Leg Pain. *Spine (Phila Pa 1976)*. 2014;39(8):678-682. doi:10.1097/BRS.0000000000000220
30. Knio ZO, VanHorn TA, O'Gara TJ. Should the EuroQol-Five Dimensions Replace the Oswestry Disability Index When Tracking Lumbar Tubular Microdecompression Outcomes? *World Neurosurg*. 2020;134:e566-e571. doi:10.1016/j.wneu.2019.10.124
31. Whynes DK, McCahon RA, Ravenscroft A, Hodgkinson V, Evley R, Hardman JG. Responsiveness of the EQ-5D Health-Related Quality-of-Life Instrument in Assessing Low Back Pain. *Value in Health*. 2013;16(1):124-132. doi:10.1016/j.jval.2012.09.003
32. Carreon L, Bratcher K, Das N, Nienhuis J, Glassman S. Estimating EQ-5D Values From the Oswestry Disability Index and Numeric Rating Scales for Back and Leg Pain. *The Spine Journal*. 2013;13:S40.
33. Pellekooren S, Ben ÂJ, Bosmans JE, et al. Can EQ-5D-3L utility values of low back pain patients be validly predicted by the Oswestry Disability Index for use in cost-effectiveness analyses? *Quality of Life Research*. 2022;31(7):2153-2165. doi:10.1007/s11136-022-03082-6
34. Poder TG, Carrier N. Predicting EQ-5D-5L Utility Scores from the Oswestry Disability Index and Roland-Morris Disability Questionnaire for Low Back Pain. *J Pain Res*. 2020;Volume 13:623-631. doi:10.2147/JPR.S236957

35. Khan KA, Madan J, Petrou S, Lamb SE. Mapping between the Roland Morris Questionnaire and Generic Preference-Based Measures. *Value in Health*. 2014;17(6):686-695. doi:10.1016/j.jval.2014.07.001
36. Madan J, Khan KA, Petrou S, Lamb SE. Can Mapping Algorithms Based on Raw Scores Overestimate QALYs Gained by Treatment? A Comparison of Mappings Between the Roland–Morris Disability Questionnaire and the EQ-5D-3L Based on Raw and Differenced Score Data. *Pharmacoeconomics*. 2017;35(5):549-559. doi:10.1007/s40273-016-0483-z
37. Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. *Health Qual Life Outcomes*. 2013;11(1):151. doi:10.1186/1477-7525-11-151
38. Petrou S, Rivero-Arias O, Dakin H, et al. The MAPS Reporting Statement for Studies Mapping onto Generic Preference-Based Outcome Measures: Explanation and Elaboration. *Pharmacoeconomics*. 2015;33(10):993-1011. doi:10.1007/s40273-015-0312-9
39. Dakin H, Abel L, Burns R, Yang Y. Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: an online database and application of the MAPS statement. *Health Qual Life Outcomes*. 2018;16(1):31. doi:10.1186/s12955-018-0857-3
40. Wailoo AJ, Hernandez-Alava M, Manca A, et al. Mapping to Estimate Health-State Utility from Non-Preference-Based Outcome Measures: An ISPOR Good Practices for Outcomes Research Task Force Report. *Value in Health*. 2017;20(1):18-27. doi:10.1016/j.jval.2016.11.006
41. Dorans NJ. Linking scores from multiple health outcome instruments. *Quality of Life Research*. 2007;16(S1):85-94. doi:10.1007/s11136-006-9155-3
42. Kuhn M. Building Predictive Models in R Using the caret Package. *J Stat Softw*. 2008;28(5). doi:10.18637/jss.v028.i05
43. Apeldoorn AT, Ostelo RW, van Helvoirt H, et al. A Randomized Controlled Trial on the Effectiveness of a Classification-Based System for Subacute and Chronic Low Back Pain. *Spine (Phila Pa 1976)*. 2012;37(16):1347-1356. doi:10.1097/BRS.0b013e31824d9f2b
44. Oosterhuis T, Ostelo RW, van Dongen JM, et al. Early rehabilitation after lumbar disc surgery is not effective or cost-effective compared to no referral: a randomised trial and economic evaluation. *J Physiother*. 2017;63(3):144-153. doi:10.1016/j.jphys.2017.05.016
45. Juch JNS, Maas ET, Ostelo RWJG, et al. Effect of Radiofrequency Denervation on Pain Intensity Among Patients With Chronic Low Back Pain. *JAMA*. 2017;318(1):68. doi:10.1001/jama.2017.7918
46. van Dongen JM, van Hooff ML, Spruit M, de Kleuver M, Ostelo RWJG. Which patient-reported factors predict referral to spinal surgery? A cohort study among 4987 chronic low back pain patients. *European Spine Journal*. 2017;26(11):2782-2788. doi:10.1007/s00586-017-5201-9
47. van Hooff ML, van Dongen JM, Coupé VM, Spruit M, Ostelo RWJG, de Kleuver M. Can patient-reported profiles avoid unnecessary referral to a spine surgeon? An observational study to further develop the Nijmegen Decision Tool for Chronic Low Back Pain. *PLoS One*. 2018;13(9):e0203518. doi:10.1371/journal.pone.0203518
48. Mutubuki EN, Luitjens MA, Maas ET, et al. Predictive factors of high societal costs among chronic low back pain patients. *European Journal of Pain*. 2020;24(2):325-337. doi:10.1002/ejp.1488
49. Eekhout I, de Vet HCW, Twisk JWR, Brand JPL, de Boer MR, Heymans MW. Missing data in a multi-item instrument were best handled by multiple imputation at the item score level. *J Clin Epidemiol*. 2014;67(3):335-342. doi:10.1016/j.jclinepi.2013.09.009
50. van Hooff ML, Spruit M, Fairbank JCT, van Limbeek J, Jacobs WCH. The Oswestry Disability Index (Version 2.1a). *Spine (Phila Pa 1976)*. 2015;40(2):E83-E90. doi:10.1097/BRS.0000000000000683
51. Yeomans S, Liebenson C. Applying outcomes management to clinical practice. *Journal of the Neuromusculoskeletal System*. 1997;5(1):1-14.

52. Hudson-Cook N, Tomes-Nicholson K, Breen A. Revised Oswestry disability questionnaire. In: Roland M, Jenner J, eds. *Back Pain: New Approaches to Rehabilitation and Education*. Manchester University Press; 1989:187-204.
53. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. *Br Med Bull*. 2010;96(1):5-21. doi:10.1093/bmb/ldq033
54. Roudijk B, Donders ART, Stalmeier PFM. Setting Dead at Zero: Applying Scale Properties to the QALY Model. *Medical Decision Making*. 2018;38(6):627-634. doi:10.1177/0272989X18765184
55. Lamers LM, McDonnell J, Stalmeier PFM, Krabbe PFM, Busschbach JJ V. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Econ*. 2006;15(10):1121-1132. doi:10.1002/hec.1124
56. Sigl T, Cieza A, Brockow T, Chatterji S, Kostanjsek N, Stucki G. Content Comparison of Low Back Pain-Specific Measures Based on the International Classification of Functioning, Disability and Health (ICF). *Clin J Pain*. 2006;22(2):147-153. doi:10.1097/01.ajp.0000155344.22064.f4
57. Cieza A, Stucki G. Content comparison of health-related quality of life (HRQOL) instruments based on the international classification of functioning, disability and health (ICF). *Quality of Life Research*. 2005;14(5):1225-1237. doi:10.1007/s11136-004-4773-0
58. World Health Organisation. International Classification of Functioning, Disability and Health (ICF). Published online 2001. Accessed December 4, 2021. https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health
59. Jolliffe I. *Principal Component Analysis*. Springer Science & Business Media; 2013.
60. R: Principal components analysis (PCA). Accessed March 24, 2022. https://personality-project.org/r/psych/help/principal.html
61. van Hout BA, Shaw JW. Mapping EQ-5D-3L to EQ-5D-5L. *Value in Health*. 2021;24(9):1285-1293. doi:10.1016/j.jval.2021.03.009
62. Osborne J. *Best Practices in Logistic Regression*. SAGE Publications; 2014.
63. Lamers L, Stalmeier P, McDonnell J, Krabbe P, van Busschbach J. 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.
64. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8(2):135-160. doi:10.1177/096228029900800204
65. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Quality of Life Research*. 2005;14(6):1523-1532. doi:10.1007/s11136-004-7713-0
66. Maas ET, Juch JN, Groeneweg JG, et al. Cost-effectiveness of minimal interventional procedures for chronic mechanical low back pain: design of four randomised controlled trials with an economic evaluation. *BMC Musculoskelet Disord*. 2012;13(1):260. doi:10.1186/1471-2474-13-260
67. Löthgren M, Zethraeus N. Definition, interpretation and calculation of cost-effectiveness acceptability curves. *Health Econ*. 2000;9(7):623-630. doi:10.1002/1099-1050(200010)9:7<623::AID-HEC539>3.0.CO;2-V
68. Palsgrove A, Patton C, King P, Gelfand J, Turcotte J. A comparison of PROMIS Global Health-Mental and legacy orthopedic outcome measures for evaluating preoperative mental health status. *J Orthop*. 2020;19:98-101. doi:10.1016/j.jor.2019.11.032
69. Chiarotto A, Ostelo RW, Boers M, Terwee CB. A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain. *J Clin Epidemiol*. 2018;95:73-93. doi:10.1016/j.jclinepi.2017.11.005

70. Thompson NR, Lapin BR, Katzan IL. Mapping PROMIS Global Health Items to EuroQol (EQ-5D) Utility Scores Using Linear and Equipercentile Equating. *Pharmacoeconomics*. 2017;35(11):1167-1176. doi:10.1007/s40273-017-0541-1
71. Bland JM, Altman DG. Statistics Notes: Some examples of regression towards the mean. *BMJ*. 1994;309(6957):780-780. doi:10.1136/bmj.309.6957.780
72. Garratt AM, Furunes H, Hellum C, et al. Evaluation of the EQ-5D-3L and 5L versions in low back pain patients. *Health Qual Life Outcomes*. 2021;19(1):155. doi:10.1186/s12955-021-01792-y
73. Brooks R. Introduction. In: Brooks R, Rabin R, de Charro F, eds, eds. *The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective: Evidence from the EuroQol BIOMED Research Programme*. Springer Netherlands; 2003:1-5.
74. van Hooff ML, van Loon J, van Limbeek J, de Kleuver M. The Nijmegen Decision Tool for Chronic Low Back Pain. Development of a Clinical Decision Tool for Secondary or Tertiary Spine Care Specialists. *PLoS One*. 2014;9(8):e104226. doi:10.1371/journal.pone.0104226
75. Apeldoorn AT, Ostelo RW, van Helvoirt H, Fritz JM, de Vet HC, van Tulder MW. The cost-effectiveness of a treatment-based classification system for low back pain: design of a randomised controlled trial and economic evaluation. *BMC Musculoskelet Disord*. 2010;11(1):58. doi:10.1186/1471-2474-11-58
76. Oosterhuis T, van Tulder M, Peul W, et al. Effectiveness and cost-effectiveness of rehabilitation after lumbar disc surgery (REALISE): design of a randomised controlled trial. *BMC Musculoskelet Disord*. 2013;14(1):124. doi:10.1186/1471-2474-14-124

Appendix I. Description of the studies from which data were retrieved

MINT Study

The MINT study assessed the effectiveness and the cost-effectiveness of adding minimal interventional procedures to a standardized exercise program, compared with a standardized exercise program alone in patients with chronic LBP. It was conducted at 16 multidisciplinary pain clinics in the Netherlands and had a randomized and observational track. Patients were eligible when meeting the criteria: pain in facet joint, sacroiliac joint, or a combination of the facet joint, sacroiliac joint, or intervertebral disk, aged 18 to 70 years, and no improvement in symptoms after conservative treatment. Exclusion criteria were pregnancy, severe psychological problems, involvement in work-related conflicts or claims; body mass index higher than 35; or anticoagulant drug therapy or coagulopathy. For more details we refer to the original publications. [45,48,66]

Nijmegen Decision Tool Study

In the Nijmegen Decision Tool Study, 47 indicators for a successful treatment outcome were assessed among chronic LBP patients to compile a decision screening tool. Patients were recruited at a Dutch orthopaedic hospital specialized in spine care, prior to their first consultation at the orthopaedic outpatient department. Patients were eligible when meeting the criteria: experienced LBP complaints for more than three months due to degenerative lumbar spine disorders (excluding trauma and tumour), had access to the internet, and were able to read and write Dutch. For more details we refer to the original publication.[46,47,74]

Apeldoorn et al. Study

The study of Apeldoorn et al. assessed the effectiveness and the cost-effectiveness of a modified version of Delitto's classification-based treatment approach compared with usual physical therapy care in patients with sub-acute and chronic LBP treated in a primary care setting in the Netherlands. Patients were eligible when meeting the criteria: LBP as the primary complaint (with or without associated leg pain), age between 18 and 65 years, current episode longer than 6 weeks, and able to read and write Dutch. Exclusion criteria were known- or suspected-specific LBP, severe radiculopathy, serious co-morbidity, and psychopathology. For more details we refer to the original publications.[43,75]

REALISE Study

The REALISE Study assessed the effectiveness and the cost-effectiveness of referral for early rehabilitation after lumbar disc surgery. This multicentre, randomised, controlled trial included LPB patients with a herniated lumbar disc postoperatively treated in a

primary care facility in the Netherlands. Patients were eligible when meeting the criteria: a herniated lumbar disc confirmed by imaging and signs of nerve root compression, aged between 18 and 70 years, and ability to fill out questionnaires in Dutch. Exclusion criteria were cauda equina syndrome, neurogenic claudication, co-morbidities of the lumbar spine, spinal surgery in the prior 12 months, contraindications to exercise therapy, pregnancy, or previous lumbar disc surgery at the same level and on the same side. For more details we refer to the original publications.[44,76]

Table following Appendix I Number of missing values by variables and included studies

	MINT Study [45,48,66] n=6,316	Nijmegen study [46,47,74] n=14,859	Apeldoorn Study [43,75] n=156	REALISE Study [44,76] n=169	Total n=21,500
Age	982	0	0	3	985
Sex	983	0	0	0	983
Education level	1195	567	0	0	1762
Living together with a partner	60	0	0	0	60
Type of low back pain	1033	0	0	0	1033
Post-surgery	0	0	0	0	0
Setting	0	0	0	0	0
NRS, mean (SD)	1055	0	0	0	1055
EQ-5D-3L					
Mobility	1055	0	1	0	1056
Self-care	1055	0	0	0	1055
Usual activities	1055	0	0	0	1055
Pain/discomfort	1055	0	0	0	1055
Anxiety/depression	1055	0	0	0	1055
ODI					
Pain intensity	1059	0	0	0	1059
Personal care	1059	0	0	0	1059
Lifting	1059	0	1	0	1060
Walking	1059	0	0	0	1059
Sitting	1059	0	0	0	1059
Standing	1059	0	0	0	1059
Sleeping	1059	0	5	0	1064
Sex life	1059	2400	8	0	3467
Social life	1059	0	0	0	1059
Traveling	1059	0	0	0	1059

NRS = Numeric Rating Scale
ODI = Oswestry Disability Index

Supplementary Table 1. Conceptual overlap between ODI and EQ-5D-3L according to the ICF categories

ICF categories		EQ-5D-3L	ODI
Component: Body functions			
b1	*Domain 1 Mental functions*		
b134	Sleep functions		+
b1343	Quality to sleep		+
b152	Emotional functions	+	
b2	*Domain 2 Sensory functions and pain*		
b280	Sensation of pain	**+**	**+**
b28013	Pain in back		+
b289	Sensation of pain, other specified and unspecified	+	
Component: Activities and Participation			
d2	*Domain 2 General tasks and demands*		
d230	Carrying out daily routine	+	
d4	*Mobility*	**+**	**+**
d4103	Sitting		+
d4104	Standing		+
d4153	Maintaining a sitting position		+
d4154	Maintaining a standing position		+
d430	Lifting and carrying objects		+
d4300	Lifting		+
d450	Walking	**+**	**+**
d4501	Walking long distances		+
d4550	Crawling		+
d498	Mobility, other specified	**+**	**+**
d5	*Domain 5 Self-care*	**+**	**+**
d510	Washing oneself	**+**	**+**
d5109	Washing oneself, unspecified		+
d540	Dressing	**+**	**+**
d5409	Dressing, unspecified		+
d7702	Sexual relationships		+
d6	*Domain 6 Domestic life*		
d6409	Doing housework, unspecified	+	
d7	*Domain 7 Interpersonal interactions and relationships*		
d7609	Family relationships, unspecified	+	
d8	*Domain 8 Major life areas*		
d839	Education, other specified and unspecified	+	

Supplementary Table 1. Conceptual overlap between ODI and EQ-5D-3L according to the ICF categories

ICF categories		EQ-5D-3L	ODI
d8509	Remunerative employment, unspecified	+	
d9	*Domain 9 Community, social and civic life*		
d9209	Recreation and leisure, unspecified	**+**	**+**
Component: Environmental factors			
e1	*Domain 1 Products and technology*		
e1101	Drugs		+
e1151	Assistive products and technology for personal use in daily living		+
e1201	Assistive products and technology for personal indoor and outdoor mobility and transportation		+

ODI = Oswestry Disability Index. ICF: International Classification of Functioning, Disability and Health. Categories refers to the units of the ICF classification (i.e., lines on the table). Several categories set a specific domain. Several domains set a specific component. Each category has a code which includes its respective component represented by a letter (e.g., b, d, e) following by a numeric code. The signal + indicates that the EQ-5D-3L and/or the ODI are linked to the ICF category. In bold are the overlap between both instruments.

Supplementary Table 2. Cross-tabulation for ODI items' responses and EQ-5D-3L responses by dimensions in the training set

ODI items		EQ-5D-3L dimensions					
		Mobility			Self-care		
	Response levels	1	2	3	1	2	3
Mobility	**0**	1817	1568	13	2477	909	12
	1	602	3643	21	2641	1612	13
	2	141	2519	44	1315	1364	25
	3	39	1684	84	718	1060	29
	4	5	767	70	210	575	57
	5	3	19	48	6	36	28
Personal care	**0**	1627	2971	15	4194	410	9
	1	668	3167	39	2117	1747	10
	2	262	2996	104	945	2400	17
	3	43	880	50	104	854	15
	4	5	177	50	6	127	99
	5	2	9	22	1	18	14
Sitting	**0**	294	739	3	821	210	5
	1	533	1873	19	1578	818	29
	2	1098	3922	59	2994	2046	39
	3	516	2731	79	1550	1736	40
	4	143	803	85	363	632	36
	5	23	132	35	61	114	15
Pain intensity	**0**	69	105	2	142	32	2
	1	349	464	4	649	167	1
	2	1263	3531	30	3326	1465	33
	3	750	4083	87	2464	2408	48
	4	162	1823	112	731	1318	48
	5	14	194	45	55	166	32
Social life	**0**	717	763	4	1261	218	5
	1	734	1891	7	1830	793	9
	2	650	2103	11	1789	972	3
	3	461	4578	94	2241	2828	64
	4	33	708	122	203	608	52
	5	12	157	42	43	137	31

ODI = Oswestry Disability Index. Number of observations = 13,087

EQ-5D-3L dimensions (Continued)								
Usual activities			Pain/discomfort			Anxiety/depression		
1	2	3	1	2	3	1	2	3
576	2644	178	126	2383	889	2452	888	58
256	3614	396	46	2556	1664	2611	1535	120
90	2194	420	27	1228	1449	1531	1045	128
51	1230	526	23	619	1165	934	735	138
15	512	315	9	266	567	422	340	80
2	20	48	1	10	59	22	26	22
768	3619	226	162	3472	979	3380	1170	63
149	3269	456	36	1974	1864	2393	1343	138
59	2589	714	22	1272	2068	1669	1487	206
11	658	304	8	306	659	436	457	80
2	76	154	4	34	194	80	105	47
1	3	29	0	4	29	14	7	12
255	729	52	82	696	258	791	230	15
293	1908	224	62	1606	757	1613	753	59
320	4210	549	57	3031	1991	3184	1739	156
95	2605	626	23	1408	1895	1840	1323	163
24	655	352	6	285	740	465	443	123
3	107	80	2	36	152	79	81	30
69	95	12	62	92	22	141	29	6
210	580	27	50	719	48	635	172	10
481	3994	349	71	3944	809	3310	1437	77
183	3982	755	35	2103	2782	2852	1886	182
42	1439	616	11	192	1894	952	943	202
5	124	124	3	12	238	82	102	69
513	940	31	113	1142	229	1238	237	9
213	2303	116	39	1762	831	1916	675	41
184	2454	126	39	1859	866	1877	844	43
65	4047	1021	31	2107	2995	2573	2291	269
13	403	447	6	162	695	297	435	131
2	67	142	4	30	177	71	87	53

Supplementary Table 3. Cross-tabulation for ODI items' responses and EQ-5D-3L responses by dimensions in the training set excluding logical inconsistencies

ODI items		EQ-5D-3L dimensions					
		Mobility			Self-care		
	Response levels	1	2	3	1	2	3
Mobility	**0**	1508	1088	0	1991	605	0
	1	467	2359	0	1887	939	0
	2	91	1450	0	836	704	1
	3	19	800	0	379	435	5
	4	0	326	21	128	204	15
	5	0	10	20	2	17	11
Personal care	**0**	1388	2224	0	3313	299	0
	1	507	1833	4	1339	1005	0
	2	164	1487	8	505	1154	0
	3	26	428	9	66	397	0
	4	0	60	15	0	46	29
	5	0	1	5	0	3	3
Sitting	**0**	265	523	0	657	130	1
	1	461	1237	6	1235	464	5
	2	911	2388	1	2176	1122	2
	3	352	1368	3	928	794	1
	4	85	442	19	195	336	15
	5	11	75	12	32	58	8
Pain intensity	**0**	63	75	1	121	18	0
	1	336	401	1	611	127	0
	2	1108	2622	3	2701	1027	5
	3	426	1522	2	1163	786	1
	4	145	1297	22	593	858	13
	5	7	116	12	34	88	13
Social life	**0**	660	625	0	1116	169	0
	1	582	1314	2	1363	533	2
	2	546	1476	3	1382	642	1
	3	289	2390	5	1299	1379	6
	4	7	180	21	53	142	13
	5	1	48	10	10	39	10

ODI = Oswestry Disability Index. Highlighted in bold are the 4,928 observations that were deleted from the training set (n=13,087) due to logical inconsistencies.

EQ-5D-3L dimensions (Continued)								
Usual activities			Pain/discomfort			Anxiety/depression		
1	2	3	1	2	3	1	2	3
523	2058	15	122	2293	181	1996	596	4
217	2579	30	42	2392	392	1879	938	9
69	1433	39	22	1113	406	952	575	14
34	726	59	13	509	297	470	323	26
9	303	35	5	218	124	201	125	21
1	12	17	0	5	25	6	13	11
684	2919	9	156	3307	149	2774	834	4
117	2189	38	29	1832	483	1553	774	17
42	1544	73	12	1109	538	919	712	28
9	414	40	6	264	193	235	217	11
1	44	30	1	17	57	22	33	20
0	1	5	0	1	5	1	0	5
229	559	0	75	674	39	638	150	0
269	1435	0	58	1504	142	1246	456	2
284	3016	0	49	2816	435	2262	1027	11
71	1652	0	19	1259	445	1084	625	14
0	387	159	3	246	297	243	260	43
0	62	36	0	31	67	31	52	15
64	75	0	60	79	0	119	20	0
203	534	1	50	688	0	590	147	1
425	3297	11	64	3669	0	2738	988	7
120	1817	13	30	1920	0	1309	635	6
38	1294	132	0	166	1298	707	714	43
3	94	38	0	8	127	41	66	28
472	810	3	110	1119	56	1098	187	0
166	1726	6	34	1684	180	1439	459	0
167	1854	4	36	1787	202	1464	561	0
45	2551	88	22	1866	796	1503	1181	0
3	144	61	2	65	141	0	153	55
0	26	33	0	9	50	0	29	30

Logical inconsistencies assumed to be as any ODI response = 0, 1, 2 or 3 and any EQ-5D = 3; and any ODI = 4 or 5 and any EQ-5D = 1. **Number of observations = 8,159**

Supplementary Table 4. Differences in transition probabilities between Adjusted and Unadjusted Ordinary Logistic Regression models

	Unadjusted OLR (N=13,087)				
ODI→3L	**Mobility**	**Self-care**	**Usual Activities**	**Pain/ discomfort**	**Anxiety/depression**
0→1	0.7794	0.5629	0.7805	0.5396	0.6092
0→2	0.1992	0.4245	0.1439	0.0177	0.3491
0→3	0.0214	0.0125	0.0756	0.4427	0.0417
1→1	0.7794	0.5629	0.7805	0.5396	0.6092
1→2	0.1992	0.4245	0.1439	0.0177	0.3491
1→3	0.0214	0.0125	0.0756	0.4427	0.0417
2→1	0.1992	0.4245	0.1439	0.4427	0.3491
2→2	0.7794	0.5629	0.7805	0.0177	0.6092
2→3	0.0214	0.0125	0.0756	0.5396	0.0417
3→1	0.1992	0.4245	0.1439	0.4427	0.3491
3→2	0.7794	0.5629	0.7805	0.0177	0.6092
3→3	0.0214	0.0125	0.0756	0.5396	0.0417
4→1	0.0214	0.0125	0.0756	0.0177	0.0417
4→2	0.1992	0.5629	0.1439	0.4427	0.6092
4→3	0.7794	0.4245	0.7805	0.5396	0.3491
5→1	0.0214	0.0125	0.0756	0.0177	0.0417
5→2	0.1992	0.5629	0.1439	0.4427	0.6092
5→3	0.7794	0.4245	0.7805	0.5396	0.3491

OLR: Ordinal Logistic Regression. The unadjusted model included one of the five EQ-5D-3L dimensions as a dependent variable and the linked ODI item as an independent variable.

Adjusted OLR (N=13,087)				
Mobility	**Self-care**	**Usual Activities**	**Pain/discomfort**	**Anxiety/depression**
0.7794	0.5629	0.7805	0.5396	0.6092
0.1993	0.4245	0.1439	0.0177	0.3491
0.0213	0.0125	0.0756	0.4427	0.0417
0.7794	0.5629	0.7805	0.5396	0.6092
0.1993	0.4245	0.1439	0.0177	0.3491
0.0213	0.0125	0.0756	0.4427	0.0417
0.1993	0.4245	0.1439	0.4427	0.3491
0.7794	0.5629	0.7805	0.0177	0.6092
0.0213	0.0125	0.0756	0.5396	0.0417
0.1993	0.4245	0.1439	0.4427	0.3491
0.7794	0.5629	0.7805	0.0177	0.6092
0.0213	0.0125	0.0756	0.5396	0.0417
0.0213	0.0125	0.0756	0.0177	0.0417
0.1993	0.5629	0.1439	0.4427	0.6092
0.7794	0.4245	0.7805	0.5396	0.3491
0.0213	0.0125	0.0756	0.0177	0.0417
0.1993	0.5629	0.1439	0.4427	0.6092
0.7794	0.4245	0.7805	0.5396	0.3491

The adjusted OLR included one of the five EQ-5D-3L dimensions as a dependent variable and all ODI items as independent variables. ODI = Oswestry Disability Index. ODI→3L represents a probability of a response in ODI is also observed in EQ-5D-3L. N = number of observations

PART IV

Appendices


8

General discussion

General discussion

Around the world policymakers face challenges with access to and equal distribution of healthcare.[1] Contrary to what one might think, this is not only a problem for low- and middle-income countries. Also in the Netherlands, being a high-income country,[2] policymakers are facing challenges to keep cost in check and allocate scarce resources.[3] Not only at the policy level, but also in clinical practice challenges exist. Healthcare workers, such as general practitioners (GPs), are dealing with high workload, and patients experience access problems and face increasing healthcare costs.[4] This results in a two-tier healthcare system where on the one hand, there are people for whom care is still available because they know their way in healthcare and have sufficient resources, and on the other hand, people who do not always have access due to insufficient resources and not knowing their way.[3,4] These problems can be expected to increase as the population ages and the number of chronically ill people rises.[3] Simply investing more money and deploying more healthcare workers is not an option. Public funds are also needed for other social ends, such as education, and, as a substantial part of Dutch population already works in the healthcare sector, extra hands are not available.[3]

Now is the time to rethink available resources critically and allocate these resources as efficiently as possible. A good starting point may be to assess large patient groups that are responsible for high resource use and cost, such as patients with musculoskeletal conditions treated in primary care. 14.6% of all contacts with a GP are related to musculoskeletal conditions,[5] and associated costs amount to 342,5 million euros, which is 8.8% of total costs made in general practice.[6] But before one can make a well-informed decision on allocating resources, one needs data on healthcare utilization and costs at a disaggregated level, such as utilization and costs specified per type of musculoskeletal condition. Due to upcoming technology, we can access large amounts of these data through electronic patient files of millions of patients. These large databases provide the opportunity to evaluate healthcare utilization within general practice, calculate healthcare cost of GP guided care and identify predictive factors for having higher healthcare cost. In clinical registration data, patient characteristics such as age and gender are often recorded, but potential predictive factors such as self-reported measures of health or health services are not recorded. These self-reported measures can be obtained from other data sources, such as clinical trial data or evaluations of health services. However, before considering a self-reported measure as possible predictive factors, it is important to assess the content validity of the measurement instrument used. To illustrate, when measuring patient satisfaction -which is not only a frequently

used outcome measure to evaluate quality of care,[7,8] but also possible predictive factor of health care utilization and thus costs, as satisfied patients seem to more benefit more from healthcare and thereby have lower healthcare utilisation[9] - the best method to measure this construct is still unclear.[8,10–12]

This is also the time to rethink the deployment of healthcare workers and the way we utilize their expertise. Especially, those healthcare workers who are experiencing high workload, such as GPs. Increasing administrative tasks, taking over tasks from secondary care facilities, and having more complex patients increase workload, while the number of GPs is declining.[13,14] This raises the question how to keep GP care accessible and maintain quality of care, while lowering GPs' workload and thereby prevent GPs from leaving the work field and attract more GPs in the future. There is a range of initiatives to relief GPs' workload for various patient populations, especially for populations with chronic disease such as diabetes or mental health conditions.[15–19] Also for patients with musculoskeletal conditions there are various initiatives to support GPs. In the Netherlands, the introduction of direct access to physiotherapy in 2006 resulted in almost three quarters of patients visiting a physiotherapist without a referral from the GP.[20] Another, more recent example is the Advanced Physiotherapy Practitioner (APP) model of care, in which APPs take over tasks from GPs in the care for patients with musculoskeletal conditions.[21,22] Despite positive findings in international publications,[23,24] it is unclear how this new model of care finds its footing in the current healthcare landscape and how this will affect healthcare utilization and associated cost within the Dutch primary care system.

Besides lowering healthcare cost and decreasing workload, maintaining good quality care is essential. Although quality of healthcare can be defined in many ways, it is generally believed that quality of health services means that health care should be effective (i.e., providing evidence-based healthcare services to those who need them), should be safe (i.e., avoiding harm to people for whom the care is intended), and people-centered (i.e., providing care that responds to individual preferences, needs and values).[25] In assessing and improving quality of care PROMs are often used.[26] One of the most widely used PROMs in assessing the effectiveness of healthcare is the EQ-5D, a preference-based measurement instrument that measures health related quality of life.[27] The utility values derived from this instrument are also used to calculate QALYs' which combine both the quality and quantity of life into a single outcome.[28] QALY's are the preferred outcome in economic evaluations that assess whether a new intervention is cost-effective (i.e., provides good value for money) compared to an alternative intervention. However, quality-of-life measurements are generally not available in electronic patient files, as these data are collected for clinical purposes.[29,30] Therefore, researchers are exploring ways to predict EQ-5D based utility values by

means of outcomes of other available PROMs.[31–33] However, no consensus has been reached on what is the best way to do so.[32,34]

Addressing rising healthcare cost and shortage of healthcare workers, while maintaining quality of care, is one of the steps towards equal access to healthcare that remains affordable. To help policymakers make well informed decisions in these issues and allocate limited resources as efficiently as possible, more insight is needed. Therefore, the objectives of this thesis were:

1. To evaluate healthcare utilization and cost of GP-guided care in patients with musculoskeletal complaints at a disaggregated level and identify predictive factors for having higher healthcare cost.
2. To assess the content validity of measurement instruments that measure patient satisfaction, as this is a frequently measured parameter of quality of care that influences healthcare utilization.
3. To evaluate the introduction of an APP model of care in Dutch general practice.
4. To explore the deployment of APPs in general practice by identifying APP patient population and evaluating APP-led health care pathways and associated cost.
5. To evaluate the predictive performance of different types of prediction modeling (i.e., linear regression analysis and response mapping approaches) to explore which method performs best in predicting EQ-5D based utility values by using the Oswestry Disability Index.

Below, the main findings of this thesis are summarized, and subsequently discussed and compared with existing literature, after which methodological considerations are discussed and implications for practice and future research are proposed.

8.1 Main findings

Part 1. Healthcare utilization and cost

In **Chapter 2** we described healthcare utilization and cost of GP-guided care in patients with musculoskeletal complaints using data of 403.719 patients and identified factors that predict higher healthcare cost. We found that the mean annual healthcare cost of GP-guided care in patients with a musculoskeletal condition was relatively low and did not differ considerably across specific conditions, except that costs were higher for low back pain with radiotherapy. Although the number of referrals varied among different type of complaints, key cost driver of GP-guided care in patients with musculoskeletal conditions was referrals to primary caregivers. The top 5% of high-cost users were

responsible for 24% of the costs. High age, being female, low social economic status, spine complaints, high number of musculoskeletal diagnoses, and high comorbidity score were found to be predictive factors in having higher healthcare cost, but only explained a negligible part of the variance in cost. Thus, it is unclear which factors do explain high healthcare cost in this population. In **Chapter 3** we performed a systematic review to assess content validity of measurement instruments that measure patient satisfaction, as this is an often-measured parameter that influences healthcare utilization. We found that all seven included PROMs used to measure patient satisfaction had insufficient content validity and the quality of this evidence was low due to shortcomings in the development of the PROMs and the lack of validation studies.

Part II. Advanced practice physiotherapy within Dutch primary care

In **Chapter 4** we evaluated the introduction of an APP model of care in Dutch primary care by conducting an explorative and interpretative qualitative study among 13 APPs and 3 GPs who were in various stages of implementing an APP care model. The results showed that implementing an APP model of care within the Dutch healthcare system is challenging. The deployment of APP is not sufficiently aligned with the core values of GPs, and GPs appear to be reluctant to hand over some control over patient care to APPs. Therefore, APPs do not appear to have ownership over the implementation, given their strong dependence on the practice, values and needs of GPs. In **Chapter 5** we discuss the findings of an explorative study in which the care for patients with musculoskeletal complaints, traditionally delivered by a GP, was delivered by APPs. This study showed that a significant part of patients who consult an APP are characterized by long-term recurrent complaints and a history of diagnostic imaging and previous treatment. During the identification of APP-led care pathways, relatively high healthcare utilization was found, which may be explained by the inclusion of patients with complex complaints and the stage of development of the role of APP. The deployment of APP showed a positive trend on health outcomes.

Part III. EQ-5D based utility values

In **Chapter 6** we evaluated the predictive ability of different linear regression models in estimating EQ-5D based utility values in low back pain patients using the Oswestry Disability Index (ODI). We developed and validated six different models using Ordinary Least Squares (OLS) and Tobit model: (1) OLS, with the total ODI score, 2) OLS, with the ODI item scores as continuous variables, 3) OLS, with the ODI item scores as ordinal variables, 4) Tobit model, with the total ODI score, 5) Tobit model, with the ODI item scores as continuous variables, 6) Tobit model, with the ODI item scores as ordinal variables. The OLS and Tobit models with continuous ODI item scores were the best performing models and showed similar probabilities of cost-effectiveness compared to the Dutch 3L value set. Results of this study suggested that the ODI can be used

to predict LBP patients' EQ-5D-3L utility values when the aim is to perform a cost-effectiveness analysis for QALYs, if utility values are missing. The models were not suitable for estimating utility values for individual patients. Literature suggests that response mapping approaches perform better than regression models when predicting utility values from a condition-specific patient-reported outcome measure because they align the scales between instruments in such a way that the distributions of their responses are matched.[34–36] Therefore, in **Chapter 7,** we developed different response mapping approaches (i.e., a non-parametric approach, a non-parametric approach excluding logical inconsistencies, and an ordinal logistic regression) and compared their predictive performance with that of the earlier developed linear regression models. The non-parametric approaches had the best predictive ability. All developed response mapping approaches were not valid for estimating individual patients' EQ-5D-3L utility values, and – depending on the approach – may considerably impact cost-utility results. In addition, the response mapping approaches did not perform better in terms of predicting individual patients' utility values than the best performing models from **Chapter 6**. Thus, based on our results response mapping approaches are not necessarily preferred over regression models for mapping PROMs to EQ-5D-3L.

8.2 Interpretation of findings & comparison with literature

Part I. Healthcare utilization and cost

During the analysis of healthcare costs of GP guided care, we found that assigned International Classification of Primary Care code (ICPC code) do not always represent a single complaint consultation. For example, patients with an ICPC code for low back pain complaints were referred to an ophthalmologist. This may indicate that during a consultation more than one health complaint was discussed, as for one consultation only one ICPC code can be used by a GP. This was also touched upon by GPs in the qualitative study (**Chapter 4**) and is in line with a publication of Salisbury et al,[37] who found that - on average - GP consultations included the discussion of 2.5 complaints across a wide range of disease areas. These complaints were not only introduced by patients but also doctors raised problems in 43% of all consultations. The ability of GPs to discuss multiple complaints within one consultation is not only due to the generalist nature of the profession but is also enabled by so-called "accumulated knowledge".[38] The mechanism of accumulated knowledge partly stems from GPs being a fixed point of contact - as Dutch patients are registered by their own GP - and being gatekeeper for secondary health facilities.[39,40] From this, it can be concluded that when assessing healthcare utilization, and thus healthcare costs, it is difficult to distinguish between consultations and allocate costs to a single complaint.

Moreover, we found that the use of clinical registration data compromised the validity of our findings. Because data were not collected for research purposes, shortcomings arise in, for example, data availability and completeness of measurements. These validity issues are also addressed in several methodological papers. [41-43] These papers emphasize that although using data from electronic health records has advantages, such as lower overhead costs, lower burden of study recruitment, and more precise estimates (i.e., higher reliability), it is important to acknowledge that challenges still need to be overcome to improve data quality and ensure sufficient validity.

Part II. Advanced practice physiotherapy within Dutch primary care

We have found that it is challenging for APPs to find footing in the current healthcare landscape due to barriers at the cultural, practical, and structural level. It may be helpful to look at factors that influenced the uptake of other healthcare workers that have been introduced in Dutch general practice the past years, such as the GP-based nurse specialist ('praktijkondersteuner huisarts'; POH) and physician assistant (PA). Recently, the Netherlands Institute for Health Services Research (NIVEL) published a study on the diversity, opportunities, and barriers of task reallocation in the Netherlands.[44] Positive factors for task reallocation from GP to GP-based nurse specialist were support of GPs (and GP organizations) and health insurers, making formal agreements at an early stage in the process, and establishing a professional association representing the interests of GP-based nurse specialists. Barriers that were identified were lack of clarity regarding task definition and scope, no robust funding structure, emerging of other new functions that might make the role of general practice-based nurse specialist redundant, and too few traineeships in general practice.

In the implementation of GP-based nurse specialist there was a strong support of stakeholders such as GPs (organizations) and health insurers. Due to an ever-expanding range of tasks and increasing workload, there was a great need among GPs for support alongside the doctor's assistant who could take over more complex care tasks.[44] This great need of GPs and support of health insurers did not emerge in the interviews among GPs and APPs (**Chapter 4)**. In addition to support from GPs and health insurers, acceptance of APP by other stakeholders, such as medical specialist and patients, is important. Shortly after publication of the qualitative paper on APP (**Chapter 4),** the Dutch Orthopedic Association (NOV) called upon their members to not participate in an upcoming research trial on task reallocation of musculoskeletal care in general practice that was initiated by the Royal Dutch Society for Physical Therapy (KNGF) and Zorgverzekeraars Nederland.[45] The NOV indicated that they recognize that direct referral by physiotherapists can lead to improved patient satisfaction and efficiency gain in the process, however, the overview that GPs currently have is very important for

the patient. This is in line with the earlier discussed study of Hjortdah[38] that describes the advantages of 'accumulated knowledge' of an individual patient. In the qualitative study (**Chapter 4**) APPs indicated the lack of accumulated knowledge in setting out healthcare pathways and therefore they had to rely on the limited amount of information patients are willing to provide in a single consultation. In addition, the NOV is concerned that when making physiotherapists gatekeeper, referral rates for secondary orthopedic care will rise, hence increasing costs. This is in line with the findings of the systematic review by Sripa et al,[46] which showed that gatekeeping by GPs is associated with lower care utilization and expenditure, and better quality of care compared with direct access. According to NOV a more sustainable solution for improving the quality of referral is to improve the lines of communication between GPs and medical specialists, and for GPs to work more closely with physiotherapists.

Another facilitator in the implementation of a GP-based nurse specialist model was the active representation of GP-based nurse specialists' interests by professional associations. According to the report of the Netherlands Institute for Health Services Research[44] professional associations of GP-based nurse specialists were actively involved in gaining a foothold of GP-based nurse specialist in general practice. In the qualitative study (**Chapter 4),** we found that APPs perceived a lack of a decisive professional association and that both the APP and physiotherapy professional associations are not yet aligned when it comes to task reallocation within musculoskeletal care. For example, the professional organisation of APP states that only specially trained physiotherapists are suitable to take on this role[47] where the KNGF uses a broader criterion in which there are fewer educational requirements, partly to reduce the outflow of physiotherapists by offering more career prospects.[48,49]

During the introduction of the GP-based nurse specialist, there was a lack of clarity about the responsibilities and tasks and what qualifications a GP-based nurse specialist should have. This created a great diversity of tasks and difference in levels within the position of GP-based nurse specialist. This was also found in the introduction of APP (**Chapter 4).** Nowadays, GP-based nurse specialist's tasks and responsibilities are clear, as they focus on prevention, guidance, monitoring and providing information and education on dealing with chronic conditions and their consequences under supervision of GPs (i.e., task delegation). Due to the broad scope of tasks and different specializations (e.g., diabetes, asthma/COPD, cardiovascular diseases, or elderly care) GPs can deploy a GP-based nurse specialist where their needs are met. Unlike GP-based nurse specialist, APP's scope is limited to the diagnostic process and setting out care pathways with a focus on patients with complex and highly complex conditions. This limits the available number of patients and minimizes referral streams. This was also

touched upon in the qualitative study (**Chapter 4)** in which GPs and APPs indicated that one GP alone doesn't have enough patients for one APP and that collaboration between GP practices is essential for establishing sufficient referral stream to get an APP-practice started. Moreover, the APP model of care is based on task reallocation, leading to less control of GPs in the care for patients that consult an APP compared to patients that consult a GP-based nurse specialist. Collaboration between GP practices was also of importance during the introduction of GP-based nurse specialists. Therefore, a financial incentive for GPs was initiated by the Ministry of Health, Welfare and Sport to promote the deployment of GP-based nurse specialists. However, little is known about the effects of this incentive.[44]

A further increase in the number of patients with chronic and mental health conditions means that GPs will need more support from GP-based nurse specialists when caring for these patient groups.[44] This contrasts with patients with musculoskeletal conditions for whom, according to a national healthcare utilization report, the number of disease episodes is slightly decreasing.[50] The Public Health Foresight Study, commissioned by the Ministry of Health, Welfare and Sport,[51] showed that by 2040 the number of disability-adjusted life years for musculoskeletal conditions in general will decrease, except for back and neck pain and osteoarthritis. This may indicate that to be successfully deployed, APPs should focus on these specific patient populations. In addition to the expected decrease in prevalence and decline in disability-adjusted life years, the number of patients consulting the GP for musculoskeletal disorders has been steadily decreasing in recent years because of Direct Access Physiotherapy, towards 71% of the patients consulting a physiotherapist without a GP referral in 2021.[20] This could be a possible explanation why GPs do not perceive the added value of reallocating musculoskeletal care in their own settings, unlike other healthcare providers, such as those working in musculoskeletal NHS clinics in England.[24]

What also may influence GPs in adopting an APP care model is the experience gained in reallocating tasks to GP-based nurse specialist. Research shows consistently that the deployment of GP-based nurse specialist does not or hardly leads to a reduction in workload. This concerns both perceived workload and objective workload.[44] This can be explained by the fact that the GP-based nurse specialist makes up only a small proportion of the total capacity within the GP setting, delegating tasks in turn brings new management tasks, and the care of chronic patients intensifies as GP-based nurse specialist performs tasks that were previously not, or less extensively, performed. However, the main reason seems to be that the time freed up is filled by the increasing demand for GP care in general.[44]

The study on task reallocation by the Netherlands Institute for Healthcare Research also reports that from a patient perspective, quality of care was maintained when deploying GP-based nurse specialist.[44] Although we found that GPs have confidence in APP's musculoskeletal expertise and recognize its added value in the GP setting, our study also found that GPs are concerned about further fragmentation of care may negatively affect quality of care. These concerns seem valid as several studies showed that continuity of care is associated with collecting accumulated knowledge, increased patient satisfaction, increased take-up of health promotion, greater adherence to medical advice and decreased use of hospital services.[38,52-54] This seem to be especially important for patients who are less able to find their own care, such as patients with a low economic status, and who come to the GP with all their concerns, including social problems.[55] Therefore, GPs are more likely to prefer APPs on a consultative basis rather than in the case-manager role.

Part III. EQ-5D based utility values

In this part we evaluated the predictive ability of different predictive modes (i.e., linear regression and non-parametric approaches) in estimating EQ-5D-3L based utility values for use in cost-effectiveness analyses among low back pain patients using the Oswestry Disability Index. Given the relatively low absolute/relative fit and poor agreement between estimated and observed utility values, regression-based models were not recommended for estimating utility values for individual patients. This is line with finding of Carreon et al.[33] who concluded that individual patients' EQ-5D-3L utility values could not validly be predicted from their ODI scores. The three response mapping approaches that we developed in **Chapter 7** did not perform better in terms of predicting individual patients' utility values than the best performing models from **Chapter 6.**

Thus, based on our results response mapping approaches are not necessarily preferred over regression models for mapping PROMs to EQ-5D-3L. This finding is in contrast with the theoretical literature that suggests that response mapping generally performs better than regression models to predict utility values.[34-36] A possible explanation for this discrepancy might be that there were not many extreme scores (i.e., a ceiling effect) making regression to the mean less likely to occur, and that a relatively small proportion of the sample had a utility value of 1.

8.3 Methodological considerations and limitations

Part I. Healthcare utilization and cost

In **Chapter 1** we evaluated healthcare utilization and cost of GP guided care in musculoskeletal complaints using clinical registration data. This kind of real-life data includes large amount of data, hence accounts for reliability issues one faces with small samples, such as imprecise effect sizes and low reproducibility of results.[56] Although these large amounts of data provide precision at a high level, challenges arise when it comes to validity. Issues with representativeness, data availability and interpretation, missing measurements and missing visits are threats to validity.[57] These issues led to several shortcomings in calculating healthcare utilization and predictive modelling in our study. Inaccurate or incomplete record keeping of ICPC codes may have led to missing patients with a musculoskeletal condition and underestimation of factors that may influence healthcare utilization. The study of Salisbury et al,[37] also showed that only 37% of problems discussed were coded in electronic medical records. Flaws in registration were also found in our data. For example, in our predictive model one of the included variables was obesity (yes/no), which was assigned based on registered ICPC code. In our sample, including 403.719 patients, only 0.3% was obese, which is a fraction of the 14.3% of Dutch population being obese according to a report of the National Institute for Public Health and the Environment.[58] In identifying prognostic factors for having higher healthcare cost, we were not able to include modifying predictive factors such as self-reported physical functioning or health related quality of life, as the clinical registration data that we used did not include these variables. When these types of variables are included in an analysis, it is important to consider the way these variables were measured. In **Chapter 3** we found that content validity of PROMs used to measure patient satisfaction was insufficient. This is in line with other studies that have assessed content validity of measurement instruments that aim to measure constructs that are also known to be a modifying predictive factor in patients with musculoskeletal conditions, such as Health related Quality of Life, [59] self-reported physical functioning,[60] and pain.[61] Including outcomes derived from measurement instruments with insufficient content validity will lead to validity issues, hence bias in the prediction model.

Part II. Advanced practice physiotherapy within Dutch primary care

In **Chapter 4** we explored the experiences and perceptions of APPs and GPs towards both the deployment and implementation of APP within Dutch primary care and found that it is difficult for APPs to carve out a place for themselves within the healthcare landscape. However, the transferability of our findings is unclear. Despite similar findings in extant literature on implementation level, comparison with international literature

is difficult given the specific Dutch context. Although we used maximum variation sampling, we were compelled to recruit GPs through convenience sampling given the limited number of GPs who were willing to participate. This meant that we failed to include GPs who were not open to implementing the APP model. This probably hinders the transferability of our findings, as we may have missed aspects of the GP perspective. However, gaining trust in APP, the need for a clear added value, reluctance to hand over control, and strongly held core values were expressed as barriers for implementation by the three participating GPs. There may also be shortcomings in the dependability of the findings. Although we collected data until no new themes were identified and we used flexible analysis, data collection and data analysis were not a fully iterative process. In addition, there is a possibility that some of the participants may have felt less free to express themselves during the interview, out of concern that they might be recognized by colleagues and stakeholders based on their specific characteristics despite being anonymized. The results of the exploratory study on the deployment of APP in different GP practices (**Chapter 5**) were difficult to interpret and generalize due to the small sample size and large amount of loss to follow-up. However, the results reflect some findings from the qualitative study, such as ambiguities during patient triage, difficulties in initiating patient flow and the importance of scaling up in a larger team.

Part III. EQ-5D based utility values

In developing the predictive models, we used EQ-5D-3L utilities instead of EQ-5D-5L utilities. This is a limitation because EQ-5D-5L is known to be more sensitive and therefore recommended in pharmacoeconomic guidelines.[62,63] Nonetheless, some countries still use the EQ-5D-3L. Therefore, we preferred using the current relatively large dataset with EQ-5D-3L utility values of nearly 20,000 patients for developing and validating the models, instead of using a relatively small dataset with EQ-5D-5L. As the performance measures in the sensitivity analysis using the EQ-5D-5L reversed cross walk were comparable with those of the EQ-5D-3L version, we expect that EQ-5D-5L values can also be validly estimated using ODI scores. Another limitation was that for assessing the performance of the developed models in a trial-based cost-effectiveness analysis setting, we only used data of two clinical trials, both of which demonstrated that the probability of the interventions being cost-effective was low regardless of the willingness-to-pay threshold. In datasets where the interventions' cost-effectiveness is less conclusive, even small differences in the probability of an intervention being cost-effective might impact the overall conclusion of a study. This can be especially of relevance in comparing the 3L version with the more sensitive 5L version of EQ-5D.

8.4 Implications for further research and practice

Part I. Healthcare utilization and cost

Our findings showed that the mean annual healthcare cost of GP-guided care in patients with a musculoskeletal condition was relatively low and did not differ considerably across specific conditions. However, since electronic health record results have several limitations that affect validity, validation of our findings using other sources and methods (e.g., economic evaluations and qualitative research) is needed. Also, Including patients/family and costs in other sectors (i.e., productivity losses) would be informative in calculating total healthcare cost. The predictors of having higher healthcare costs should be validated using other datasets and the model should preferably be complemented with PROMs. However, when PROMs are included as an independent variable, they should have sufficient content validity.

Part II. Advanced practice physiotherapy within Dutch primary care

Our findings demonstrated that implementing an APP model of care is challenging, in part, because the deployment of APP does not sufficiently align with the core values of GPs, and GPs appear reluctant to hand over control of elements of patient care to APPs. In addition, APPs do not appear to have ownership over the implementation, given their strong dependence on the practice, values and needs of GPs. This means that much work still needs to be done to better embed APP in the primary care landscape in the Netherlands. Based on our findings there are several overarching factors that can facilitate the deployment and implementation of APP. These include a clearly formulated added value of APP that matches the need and demand of GPs, setting up partnerships in line with the future vision of GP care, visibility of APP among the stakeholders involved, reimbursement for APP from basic health insurance, and authority to independently refer to second care facilities. Within the profession of APP, clarity about the role and competences of APP, clarity about conforming to the NHG standards, support for APP from the APP professional association, and training at an appropriate level with sufficient depth, contact hours and practical training are facilitating factors. At the level of the individual APP, setting up a joint consultation with the GP, working in a setting independent of one's own physiotherapy practice, scaling up practice with multiple APPs and GPs, and a good relationship with physiotherapy practices in the region are factors that favor setting up an APP practice. Qualitative studies with an accent on the GP perspective can be beneficial to overcome barriers in further implementation.

Our findings also showed that a significant part of patients who consult APP

are characterized by long-term recurrent complaints and a history of diagnostic imaging and previous treatment. In addition, relatively high care utilization was found during the identification of APP-led care pathways. Given the limited number of participating APPs and low inclusion rates, results should be interpreted with caution. Our findings can be seen as a starting point for further research on the deployment of APP. In follow-up research, a framework on developing and evaluating complex interventions[64] could guide possible next steps. This framework offers the possibility of answering questions around complex interventions from different perspectives and methods, identifying the most relevant outcomes in the different development phases and alternating research between the different development phases.

Part III. EQ-5D based utility values

Our findings suggest that predictive modelling can be used to estimate utility values from disease-specific measures, such as the ODI amongst LBP patients, when assessing incremental costs per QALY gained (as part of a cost-effectiveness analysis) or differences in utilities between groups. This is helpful for assessing cost-effectiveness in trials that did not directly measure utilities. Given the relatively large root mean square error (RMSE, i.e., low absolute fit of the models) and the relatively low r-square value (i.e., low relative fit) it is strongly discouraged to use the developed models for estimating utility values of individual patients. Further research is needed to validate the models in order to 1) assess whether these models yield comparable results, to those we have found, in other empirical datasets on LBP interventions, especially in analysis on interventions that are expected not to be more conclusive in their cost-effectiveness, and 2) to improve their generalizability among different LBP patients by external validation in another sample of LBP patients than the one used in this study. The developed mapping approaches did not perform better than regression-based models and are therefore not recommended for use in economic evaluations. Further research comparing response mapping approaches and regression-based models for estimating EQ-5D utility values is needed.

8.5 Concluding remarks

This thesis explored ways to contribute to constraining healthcare costs and rethink the use of healthcare professionals and how we use their expertise for musculoskeletal problems in primary care. Emerging developments were used, such as the use of electronic health records, the APP model of care, and response

mapping approaches. Findings showed, amongst others, that it is challenging to allocate healthcare utilization and associated costs to a single complaint, identify predictive factors for having higher healthcare costs, and introduce APP as a new model of care. But most important, findings showed that primary care is a complex and dynamic healthcare landscape in which it is challenging to decrease GPs workload while maintaining quality of care, and not lose sight of the elements of personalized care and the needs of individual patients, such as giving people choice and control over how their care is planned and delivered.

References

1. World Health Organisation. Improving service access and quality. Accessed February 14, 2023. https://www.who.int/activities/improving-service-access-and-quality
2. World Health Organization. Income level. https://www.who.int/data/gho/indicator-metadata-registry/imr-details/193. Accessed February 14, 2023. https://www.who.int/data/gho/indicator-metadata-registry/imr-details/193
3. Government of the Netherlands. Integraal Zorgakkoord: "Samen Werken Aan Gezonde Zorg." Accessed February 14, 2023. https://www.rijksoverheid.nl/documenten/rapporten/2022/09/16/integraal-zorgakkoord-samen-werken-aan-gezonde-zorg
4. Dutch Healthcare Authority (NZa). De Zorg Heeft Een Kritisch Punt Bereikt. Stand van de Zorg 2022.; 2022. Accessed February 14, 2023. https://magazines.nza.nl/nza-magazines/2022/04/1-de-zorg-heeft-een-kritisch-punt-bereikt
5. Flinterman L, Groenewegen P, Verheij R. Zorglandschap En Zorggebruik in Een Veranderende Eerste Lijn.; 2018.
6. National Institute for Public Health and the Environment. Kosten van ziekten. Accessed February 14, 2023. https://statline.rivm.nl/#/RIVM/nl/dataset/50050NED/table?ts=1676300847815
7. Desmeules F, Roy JS, MacDermid JC, Champagne F, Hinse O, Woodhouse LJ. Advanced practice physiotherapy in patients with musculoskeletal disorders: a systematic review. BMC Musculoskelet Disord. 2012;13(1):107. doi:10.1186/1471-2474-13-107
8. Fennelly O, Blake C, Desmeules F, Stokes D, Cunningham C. Patient-reported outcome measures in advanced musculoskeletal physiotherapy practice: a systematic review. Musculoskeletal Care. 2018;16(1):188-208. doi:10.1002/msc.1200
9. Fitzpatrick R. The assessment of patient satisfaction. In: Jenkinson J, ed. Assessment and Evaluation of Health and Medical Care. Open University Press; 1997:85-102.
10. Batbaatar E, Dorjdagva J, Luvsannyam A, Amenta P. Conceptualisation of patient satisfaction: a systematic narrative literature review. Perspect Public Health. 2015;135(5):243-250. doi:10.1177/1757913915594196
11. Hush JM, Cameron K, Mackey M. Patient Satisfaction With Musculoskeletal Physical Therapy Care: A Systematic Review. Phys Ther. 2011;91(1):25-36. doi:10.2522/ptj.20100061
12. Pascoe GC. Patient satisfaction in primary health care: A literature review and analysis. Eval Program Plann. 1983;6(3-4):185-210. doi:10.1016/0149-7189(83)90002-2
13. Schäfer W, van den Berg M, Groenewegen P. De werkbelasting van huisartsen in internationaal perspectief. Huisarts Wet . 2016;59:94-101.
14. van der Horst H, Bindels P, Assendelft P, et al. Hoogste tijd voor minder patiënten per huisarts. Huisarts Wet. 2018;61(3):30-31. doi:10.1007/s12445-018-0056-9
15. Landelijke Huisartsen vereniging (LHV). Meer tijd voor de patiënt. Accessed February 14, 2023. https://www.lhv.nl/thema/praktijkzaken/meer-tijd-voor-de-patient/
16. Halcomb EJ, McInnes S, Patterson C, Moxham L. Nurse-delivered interventions for mental health in primary care: a systematic review of randomized controlled trials. Fam Pract. 2019;36(1):64-71. doi:10.1093/fampra/cmy101
17. Baker E, Gwernan-Jones R, Britten N, et al. Refining a model of collaborative care for people with a diagnosis of bipolar, schizophrenia or other psychoses in England: a qualitative formative evaluation. BMC Psychiatry. 2019;19(1):7. doi:10.1186/s12888-018-1997-z

18. Bennett M, Walters K, Drennan V, Buszewicz M. Structured Pro-Active Care for Chronic Depression by Practice Nurses in Primary Care: A Qualitative Evaluation. PLoS One. 2013;8(9):e75810. doi:10.1371/journal.pone.0075810
19. Laurant M, van der Biezen M, Wijers N, Watananirun K, Kontopantelis E, van Vught AJ. Nurses as substitutes for doctors in primary care. Cochrane Database of Systematic Reviews. 2018;2019(2). doi:10.1002/14651858.CD001271.pub3
20. Veldkamp R, Meijer W. Zorg Door de Fysiotherapeut. Nivel Zorgregistraties Eerste Lijn, Jaarcijfers 2021 En Trendcijfers 2018-2021.; 2022.
21. Kersten P, McPherson K, Lattimer V, George S, Breton A, Ellis B. Physiotherapy extended scope of practice – who is doing what and why? Physiotherapy. 2007;93(4):235-242. doi:10.1016/j.physio.2007.02.007
22. Tawiah AK, Desmeules F, Finucane L, et al. Advanced practice in physiotherapy: a global survey. Physiotherapy. 2021;113:168-176. doi:10.1016/j.physio.2021.01.001
23. Lafrance S, Demont A, Thavorn K, Fernandes J, Santaguida C, Desmeules F. Economic evaluation of advanced practice physiotherapy models of care: a systematic review with meta-analyses. BMC Health Serv Res. 2021;21(1):1214. doi:10.1186/s12913-021-07221-6
24. Vedanayagam M, Buzak M, Reid D, Saywell N. Advanced practice physiotherapists are effective in the management of musculoskeletal disorders: a systematic review of systematic reviews. Physiotherapy. 2021;113:116-130. doi:10.1016/j.physio.2021.08.005
25. World Health Organization. Quality of care. Accessed February 14, 2023. https://www.who.int/health-topics/quality-of-care#tab=tab_1
26. Øvretveit J, Zubkoff L, Nelson EC, Frampton S, Knudsen JL, Zimlichman E. Using patient-reported outcome measurement to improve patient care. International Journal for Quality in Health Care. 2017;29(6):874-879. doi:10.1093/intqhc/mzx108
27. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy (New York). 1990;16(3):199-208. doi:10.1016/0168-8510(90)90421-9
28. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. Measuring and Valuing Health Benefits for Economic Evaluation. Vol 1. Oxford University Press; 2016. doi:10.1093/med/9780198725923.001.0001
29. Gianola S, Frigerio P, Agostini M, et al. Completeness of Outcomes Description Reported in Low Back Pain Rehabilitation Interventions: A Survey of 185 Randomized Trials. Physiotherapy Canada. 2016;68(3):267-274. doi:10.3138/ptc.2015-30
30. Fatoye F, Gebrye T, Odeyemi I. Real-world incidence and prevalence of low back pain using routinely collected data. Rheumatol Int. 2019;39(4):619-626. doi:10.1007/s00296-019-04273-0
31. Hernández Alava M, Wailoo A, Pudney S, Gray L, Manca A. Mapping clinical outcomes to generic preference-based outcome measures: development and comparison of methods. Health Technol Assess (Rockv). 2020;24(34):1-68. doi:10.3310/hta24340
32. Longworth L, Rowen D. Mapping to Obtain EQ-5D Utility Values for Use in NICE Health Technology Assessments. Value in Health. 2013;16(1):202-210. doi:10.1016/j.jval.2012.10.010
33. Carreon LY, Bratcher KR, Das N, Nienhuis JB, Glassman SD. Estimating EQ-5D Values From the Oswestry Disability Index and Numeric Rating Scales for Back and Leg Pain. Spine (Phila Pa 1976). 2014;39(8):678-682. doi:10.1097/BRS.0000000000000220
34. Fayers PM, Hays RD. Should Linking Replace Regression When Mapping from Profile-Based Measures to Preference-Based Measures? Value in Health. 2014;17(2):261-265. doi:10.1016/j.jval.2013.12.002
35. Dorans NJ. Linking scores from multiple health outcome instruments. Quality of Life Research. 2007;16(S1):85-94. doi:10.1007/s11136-006-9155-3

36. Wailoo AJ, Hernandez-Alava M, Manca A, et al. Mapping to Estimate Health-State Utility from Non-Preference-Based Outcome Measures: An ISPOR Good Practices for Outcomes Research Task Force Report. Value in Health. 2017;20(1):18-27. doi:10.1016/j.jval.2016.11.006
37. Salisbury C, Procter S, Stewart K, et al. The content of general practice consultations: cross-sectional study based on video recordings. British Journal of General Practice. 2013;63(616):e751-e759. doi:10.3399/bjgp13X674431
38. Hjortdahl P. Continuity of care: general practitioners' knowledge about, and sense of responsibility toward their patients. Fam Pract . 1992;9:3-8.
39. Gérvas J, Pérez Fernández M. Western European best practice in primary healthcare. European Journal of General Practice. 2006;12(1):30-33. doi:10.1080/13814780600757187
40. Groenewegen P, Schäfer W, Schellevis F, Boerma W. Kernwaarden van Nederlandse huisartsen in internationaal perspectief. Huisarts Wet. 2020;63(7):57-61. doi:10.1007/s12445-020-0760-0
41. Sollie A, Numans M, Helsper C. Zijn de gegevens uit uw HIS een goudmijn? Huisarts Wet. 2018;61(2):14-17. doi:10.1007/s12445-018-0005-7
42. Thygesen LC, Ersbøll AK. When the entire population is the sample: strengths and limitations in register-based epidemiology. Eur J Epidemiol. 2014;29(8):551-558. doi:10.1007/s10654-013-9873-0
43. Gianfrancesco MA, Goldstein ND. A narrative review on the validity of electronic health record-based research in epidemiology. BMC Med Res Methodol. 2021;21(1):234. doi:10.1186/s12874-021-01416-5
44. Tuyl L, van Vis E, Bosmans M, Friele R, Batenburg R. Visies Op Taakherschikking: Een Inventariserend Onderzoek Naar de Diversiteit, Kansen En Belemmeringen van Taakherschikking in Nederland.; 2020.
45. Nederlandse Orthopeadische Vereniging (NOV). Standpunt NOV t.a.v. Direct Door Kunnen Wijzen Naar Medisch Specialisten Door Fysiotherapeuten. Accessed February 14, 2023. https://www.orthopeden.org/kwaliteit/kwaliteitsbeleid/standpunten-en-adviezen
46. Sripa P, Hayhoe B, Garg P, Majeed A, Greenfield G. Impact of GP gatekeeping on quality of care, and health outcomes, use, and expenditure: a systematic review. British Journal of General Practice. 2019;69(682):e294-e303. doi:10.3399/bjgp19X702209
47. Nederlandse Vereniging voor Extended Scope Specialisten. Beroepscompetentieprofiel Extended Scope Specialist Advanced Clinical Practitioner. Accessed February 14, 2023. https://nves.nl/wp-content/uploads/2022/02/20211213_bp-es-2.0.pdf
48. Koninklijk Nederlands Genootschap voor Fysiotherapie (KNGF). KNGF aan Tweede Kamer: "Werk aan toegankelijkheid en een aantrekkelijk vak." Accessed February 14, 2023. https://www.kngf.nl/actueel/2022/mei/werk-aan-toegankelijkheid-fysiotherapie-en-investeer-in-een-aantrekkelijk-vak.html
49. Koninklijk Nederlands Genootschap voor Fysiotherapie (KNGF). KNGF Beroepskolom. Accessed February 14, 2023. https://www.kngf.nl/binaries/content/assets/kngf/onbeveiligd/vak-en-kwaliteit/beroepskolom/kngf-beroepskolom-2021.pdf
50. Heins M, Bes J, Weesie Y, et al. Zorg Door de Huisarts. Nivel Zorgregistraties Eerste Lijn: Jaarcijfers 2021 En Trendcijfers 2017-2021.; 2022.
51. Ministry of Health, Welfare and Sports. Volksgezondheid Toekomst Verkenning 2018. Accessed February 14, 2023. https://www.vtv2018.nl/aandoeningen
52. Barker I, Steventon A, Deeny SR. Association between continuity of care in general practice and hospital admissions for ambulatory care sensitive conditions: cross sectional study of routinely collected, person level data. BMJ. Published online February 1, 2017:j84. doi:10.1136/bmj.j84

53. Cabana M, Jee S. Does continuity of care improve patient outcomes? J Fam Pract . 2004;53:974-980.
54. Chen CC, Cheng SH. Continuity of care and changes in medication adherence among patients with newly diagnosed diabetes. Am J Manag Care. 2016;22(2):136-142.
55. Eusman J, van Langen P, de Vroome M. Recept Voor Hulpverlening. Een Onderzoek Naar de Behoeften van Bewoners van Achterstandswijken Ten Aanzien van de Huisartsenzorg Op Sociaal-Maatschappelijk Gebied.; 2011.
56. Button KS, Ioannidis JPA, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013;14(5):365-376. doi:10.1038/nrn3475
57. Gianfrancesco MA, Goldstein ND. A narrative review on the validity of electronic health record-based research in epidemiology. BMC Med Res Methodol. 2021;21(1):234. doi:10.1186/s12874-021-01416-5
58. National Institute for Public Health and the Environment. Cijfers en feiten overgewicht. Accessed February 14, 2023. https://www.loketgezondleven.nl/gezondheidsthema/overgewicht/cijfers-en-feiten-overgewicht
59. Ghai V, Subramanian V, Jan H, Doumouchtsis SK. A systematic review highlighting poor quality of evidence for content validity of quality of life instruments in female chronic pelvic pain. J Clin Epidemiol. 2022;149:1-11. doi:10.1016/j.jclinepi.2022.04.016
60. Chiarotto A, Ostelo RW, Boers M, Terwee CB. A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain. J Clin Epidemiol. 2018;95:73-93. doi:10.1016/j.jclinepi.2017.11.005
61. Hoglund LT, Scalzitti DA, Bolgla LA, Jayaseelan DJ, Wainwright SF. Patient-Reported Outcome Measures for Adults and Adolescents with Patellofemoral Pain: A Systematic Review of Content Validity and Feasibility Using the COSMIN Methodology. Journal of Orthopaedic & Sports Physical Therapy. 2023;53(1):23-39. doi:10.2519/jospt.2022.11317
62. National Institute for Health and Care Excellence. Guide to the Methods of Technology Appraisal.; 2013. Accessed February 14, 2023. https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l
63. Hakkaart-van Roijen L, van der Linden N, Bouwmans C, Kanters T, Tan S. Kostenhandleiding: Methodologie van Kostenonderzoek En Referentieprijzen Voor Economische Evaluaties in de Gezondheidszorg.; 2015. Accessed February 14, 2023. https://www.zorginstituutnederland.nl/over-ons/publicaties/publicatie/2016/02/29/richtlijn-voor-het-uitvoeren-van-economische-evaluaties-in-de-gezondheidszorg
64. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ. Published online September 30, 2021:n2061. doi:10.1136/bmj.n2061


9

Summary

Background

Worldwide policy makers are challenged to account for rising health care costs and increased healthcare demand. Also, in the Netherlands there is a growing concern how to maintain high-quality and accessible care while keeping costs in check. Access to care is under pressure as the demand for care is rising fast, due to an aging population and an increasing number of chronically ill people. Not only at the policy level, but also in clinical practice challenges exist. The workload in the healthcare sector is high, causing health workers, such as general practitioners (GPs), to leave this sector. To keep costs in check available resources need be allocated as efficiently as possible. A good starting point for evaluating healthcare costs may be assessing large patient groups that are responsible for high resource use and costs, such as patients with musculoskeletal conditions treated in general practice. Another point may be identifying prognostic factors for higher healthcare costs. Besides lowering costs, it is also of importance to keep GP care accessible by lowering GPs' workload. One of the ways to address GPs' high workload is task reallocation. Internationally, positive effects have been found for an Advanced Physiotherapy Practitioner (APP) model of care, in which APPs take over tasks from a physician in the care for patients with musculoskeletal conditions. This model of care could potentially be of value in reducing the workload of Dutch GPs and keeping GP care accessible. Besides lowering healthcare cost and decreasing GPs' workload maintaining good quality care is essential. One of the most widely used Patient Reported Outcome Measures (PROMs) in assessing quality of healthcare is the EQ-5D, a preference-based measurement instrument that measures health related quality of life and is used to estimate utility values that represent the preferences of the general population of a country for given health states. These utility values are needed for estimating Quality-Adjusted Life-Years (QALYs) in cost effectiveness analysis. However, quality-of-life measurements are generally not available when data are collected for clinical purposes, such as data from GP electronic medical records. Therefore, researchers are exploring ways to estimate EQ-5D based utility values by means of outcomes on other available health related outcome measures. This thesis aimed to explore some of the challenges in Dutch primary care by evaluating 1) healthcare utilization and associated cost of GP-guided care in patients with musculoskeletal complaints, 2) the introduction of an APP model of care, and 3) different approaches to estimate missing EQ-5D based utility values.

Part I. Healthcare utilization and costs of musculoskeletal complaints

Chapter 2 described healthcare utilization and costs of GP-guided care in patients with musculoskeletal complaints using data from electronic medical records of 2,118,603 adult patients treated in general practice, which is 13% of all adult patients registered with a GP practice in the Netherlands. A General Linear Model was additionally developed to identify predictive factors for having higher healthcare costs. The model was internally validated using k-fold cross-validation and prognostic accuracy was

assessed using R^2 and root-mean-square error (RMSE). In total, 403,719 patients were included, of whom 92% only received a single consultation. The number of referrals varied widely across the different types of complaints. Total annual healthcare cost amounted to €39,180,531, of which the key cost driver was referrals. Referrals to primary care providers accounted for the largest part of referral-related cost. For all musculoskeletal conditions combined, mean annual healthcare cost per patient were €97 (SEM=€0.18). High age, being female, low social economic status, spine complaints, high number of musculoskeletal diagnoses, and a high comorbidity score were predictive of having higher healthcare cost and explained 0.7% of the variance. This study showed that mean annual healthcare cost of GP-guided care in patients with musculoskeletal conditions were relatively low and did not differ considerably across conditions. Investigated predictive factors explained a negligible part of the variance in cost. Thus, it is unclear which factors do explain high healthcare cost in this population. Future studies should also include costs related to productivity losses and informal care.

Chapter 3 evaluated the content validity of patient-reported outcome measurement (PROM) instruments used in patients with musculoskeletal complaints treated in primary care to assess satisfaction. A systematic review was performed in which a literature search was undertaken in MEDLINE, EMBASE and CINAHL was undertaken (up to January 2020) to identify studies of the development or evaluation of content validity of a PROM aimed to assess patient satisfaction. A PROM was considered eligible if it aimed to measure satisfaction with care in patients with musculoskeletal complaints. Two independent reviewers performed study selection, quality assessment, and data extraction. Evaluation of content validity of the included PROMs was performed according to COSMIN guidance, which includes the evaluation of the quality of a PROM development, the quality of content validity studies, the content of the PROMs, and rating the quality of evidence with a modified GRADE approach. Seven PROMs were identified. Their quality of development was inadequate. No studies evaluating content validity of the satisfaction PROMs were retrieved. The content validity of the satisfaction PROMs was insufficient and supported by very low-quality evidence. In measuring patient satisfaction among patients with musculoskeletal complaints treated in primary care, none of the identified PROMs had adequate content validity. Future studies should address relevance, comprehensiveness, and comprehensibility of PROMs used to measure satisfaction, and emphasise patient involvement during the development of new instruments.

Part II. Advanced practice physiotherapy within Dutch primary care

Chapter 4 explored the experiences and perceptions of APPs and GPs with respect to implementing APP within Dutch primary care. This explorative and interpretive qualitative study included 12 APPs and 3 GPs who were in various stages of implementing

an APP model of care. Semi-structured interviews were conducted between January and March 2021. The topic list was based on existing literature, the personal input of researchers, and the Constellation Approach framework. Data were analysed using a thematic inductive approach. Four main themes emerged from the data; 1) Both GPs' trust in APP and a clear added value of APP are critical for starting implementation, 2) APPs need continuous support from GPs, 3) APPs believe that their position needs strengthening, and 4) Implementation of the APP model creates tension over ownership. These four themes highlight the perceived difficulties in gaining trust, lack of clarity over the added value of APP, ambiguity over APPs' professional profile and positioning, a need on behalf of GPs to maintain authority, lack of reimbursement structure, and the struggle APPs face to strike a balance with current care. This study demonstrates that implementing an APP model of care is challenging, in part, because the deployment of APP does not sufficiently align with the core values of GPs, while GPs appear reluctant to hand over control of elements of patient care to APPs. APPs do not appear to have ownership over the implementation, given their strong dependence on the practice, values and needs of GPs.

Chapter 5 reports the findings of an explorative study in which the care for patients with musculoskeletal complaints, traditionally delivered by a GP, was delivered by APPs. Patients were included at four different practices between December 2020 and December 2021. Data were retrieved trough clinical registration forms and web-based questionnaires at baseline, 3-, and 6-months follow-up. Costs for usual care pathways (i.e., GP-led care pathways) were calculated using data from electronic patient records that involved data recorded by GPs as part of clinical patient care. A total of 109 patients were included and the most common condition was shoulder pain (41%). In more than half of the patients, the duration of complaints was longer than 6 months and 43% of patients had a recurrent complaint for which more than 80% had previously consulted a healthcare provider. A positive trend was seen for almost all health-related outcome measures at 3- and 6-month follow-up. Almost one-third of all APP-led pathways included a referral to a secondary care facility. Of all patients included, 71% were referred to physiotherapy, making it the most common referral. The mean costs per APP-led care pathway were €486 (SD 209) for a follow-up period of 6 months and patient-reported health costs averaged €2901 (SD=€6824) and €2729 (SD=€5715), at 3- and 6-month follow-up, respectively. The mean costs of GP-led care pathway were €97 (SD=€117) per year. A significant part of patients who consult APP are characterized by long-term recurrent complaints and a history of diagnostic imaging and previous treatment. During the identification of APP-led care pathways, relatively high care utilization was found, which may be explained by the inclusion of patients with complex complaints and the stage of development of the role of APP. The use of APP

shows a positive trend on health outcomes. Given the limited number of participating APPs and low inclusion rates, results should be interpreted with caution. Our findings can be seen as a starting point for further research on the deployment of APP. In follow-up research, a framework on developing and evaluating complex interventions could guide possible future steps in implementation.

Part III. Estimating EQ-5D-3L based utility values

Chapter 6 assesses whether regression modelling can be used to predict EQ-5D-3L utility values from the Oswestry Disability Index (ODI) in low back pain patients for use in cost effectiveness analysis. Individual patient data from 18,692 patients with low back pain were split into a training and a validation group. A total of six different regression models were developed: 1) Ordinary Least Squares (OLS) regression, with total ODI score, 2) OLS, with ODI item scores as continuous variables, 3) OLS, with ODI item scores as ordinal variables, 4) Tobit model, with total ODI score, 5) Tobit model, with ODI item scores as continuous variables, 6) Tobit model, with ODI item scores as ordinal variables. EQ-5D-3L utilities of were included as independent variable and ODI scores as independent variable. The predictive ability of the models was assessed by the explained variance (R^2) and Root Mean Squared Error (RMSE). The potential impact of using predicted, rather than measured, EQ-5D utilities on cost-effectiveness outcomes was evaluated in two empirical cost-effectiveness analyses. All models had a similar R^2(range: 45-52%) and RMSE (range: 0.21-0.22). The two best-performing models (i.e., models 2 and 5) showed relatively similar predicted probabilities of cost-effectiveness at different willingness-to-pay values compared to the observed values. For example, the difference in probabilities ranged from 2% to 5% at a willingness-to-pay value of €50,000/QALY. The results suggest that the ODI can be validly used to predict EQ-5D-3L utilities and QALYs of low back pain patients for use in cost-effectiveness analyses when EQ-5D utilities are missing. Additional research is needed to validate the models to verify whether these models yield similar results in other empirical datasets on low back pain interventions, to improve the generalizability of the estimated models, and to compare the performance of predictive models with a response mapping approach for estimating utilities. Meanwhile, researchers can use the developed models in their cost-effectiveness analysis when utilities are missing.

Chapter 7 aimed to develop and validate approaches for mapping Oswestry Disability Index (ODI) responses to EQ- 5D-3L utility values, and subsequently evaluate the impact of using mapped utility values on cost-utility results in comparison with the regression models developed in *Chapter 6*. Three response mapping approaches were developed in a random sample of 70% of 18,692 patients with low back pain: non-parametric approach (Non-p), non-parametric approach excluding logical inconsistencies (Non-peLI), and ordinal logistic regression (OLR). Performance was assessed in the remaining

30% using R-square (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). To evaluate whether MAEs and their 95% limits of agreement (95%LA) were clinically relevant, a minimally clinically important difference (MCID) of 0.074 was used. Probabilities of cost-effectiveness estimated using observed and mapped utility values were compared in two economic evaluations. The Non-p performed best (R2=0.43; RMSE=0.22; MAE=0.03; 95%LA=-0.40;0.47) compared to the Non-peLI (R2=0.07; RMSE=0.29; MAE=-0.15; 95%LA=-0.63;0.34), and ORL (R2=0.22; RMSE=0.26; MAE=0.02; 95%LA=-0.49;0.53). MAEs were lower than the MCID for the Non-p and OLR, but not for the Non-peLI. Differences in probabilities of cost-effectiveness ranged from 1-4% (Non-p), 0.1-9% (Non-peLI), and 0.1-20% (OLR). Results suggest that the developed response mapping approaches are not valid for estimating individual patients' EQ-5D-3L utility values, and – depending on the approach – may considerably impact cost-utility results. The developed approaches did not perform better than previously published regression-based models and are, therefore, not recommended for use in economic evaluations.

Discussion

In **Chapter 8**, the main findings of this thesis were discussed and interpreted based on existing literature. In conclusion, most patients with a musculoskeletal condition received a single GP consultation and the mean annual healthcare cost of GP-guided care was relatively low. Referral rates varied across different complaints and were in line with recommendations of clinical guidelines for GPs in the Netherlands. This suggests that GPs largely adhere to these guidelines and underscores the importance of guideline development and implementation. A restricted number of available predictive factors in combination with a relatively low level of variation in both the predictive factors as well as the outcome variable (i.e., healthcare costs) led to the low predictive performance of the regression model with statistically significant, but relatively small betas. However, in a large population, the budget impact can still be high. Researchers need to be aware of validity issues and limitations of using data that was not collected for research purposes, such as clinical registration data. Another conclusion was that the implementation of an APP model in Dutch primary care has proven to be challenging since APPs do not appear to have ownership over the implementation due to their strong dependence on the practice, values and needs of GPs. The implementation of an APP model of care in general practice would benefit from support of GPs and other stakeholders, a clearly defined role, standardisation of process and working methods, availability of training at an appropriate level, alignment with GP core values, and a clear added value of APP compared to already existing collaborations between GPs and physiotherapists. A

significant part of patients who consult APP are characterized by long-term recurrent complaints and a history of diagnostic imaging and previous treatment. During the identification of APP-led care pathways, relatively high care utilization was found, which may be explained by the inclusion of patients with complex complaints and the stage of development of the role of APP. Last, when EQ-5D measures are missing in studies that investigate interventions on low back pain, the condition-specific measure ODI can be used to predict EQ 5D based utility values for use in cost-effectiveness analysis, since the bias surrounding the predicted utility values is likely to be similar in the intervention and control groups, thereby not affecting incremental QALYs. This, however, is not recommended in estimating utility values for individual patients, given the low fit of regression models and poor agreement between estimated and observed utility values. In estimating EQ 5D based utility values, regression modelling is preferred over response mapping approaches, as regression models outperformed response mapping approaches and are easier to implement. Although estimating EQ 5D based utility values for use in cost-effectiveness analysis is feasible, it is strongly recommended to include health related quality of life measures, such as the EQ 5D, in the design of clinical trials when the aim is to perform a cost-effectiveness analysis.


10

Summary in Dutch | Samenvatting

Achtergrond

Wereldwijd worden beleidsmakers uitgedaagd om oplossingen te zoeken voor stijgende kosten van de gezondheidszorg en de toegenomen vraag naar zorg. Ook in Nederland is er een groeiende bezorgdheid hoe kwalitatief hoogwaardige en toegankelijke zorg te behouden en tegelijkertijd de kosten in de hand te houden. De toegang tot zorg staat onder druk nu de vraag naar zorg snel toeneemt door de vergrijzing en een toenemend aantal chronisch zieken. Niet alleen op beleidsniveau, maar ook in de klinische praktijk zijn er uitdagingen. De werkdruk in de gezondheidszorg is hoog, waardoor zorgmedewerkers, zoals huisartsen, de zorgsector verlaten. Om de kosten in de hand te houden moeten beschikbare middelen zo efficiënt mogelijk worden toegewezen. Een goed uitgangspunt voor de evaluatie van zorgkosten is om grote patiëntengroepen te beoordelen die verantwoordelijk zijn voor een hoog zorggebruik en hoge kosten, zoals patiënten met aandoeningen van het bewegingsapparaat die in de huisartsenpraktijk worden behandeld, en om prognostische factoren voor het hebben van hogere zorgkosten vast te stellen. Naast het verlagen van de kosten is het ook van belang de huisartsenzorg toegankelijk te houden door de werkdruk van de huisartsen te verlagen. Een van de mogelijke oplossingen daarvoor is taakherschikking, waarbij een andere zorgverlener taken overneemt die normaliter door een huisarts worden uitgevoerd. Internationaal zijn positieve effecten gevonden voor het inzetten van Advanced Physiotherapy Practitioners (APPs) waarbij APPs taken van een arts overnemen in de zorg voor patiënten met aandoeningen aan het houdings- en bewegingsapparaat. Wellicht dat dit zorgmodel ook van waarde kan zijn voor het verlagen van de werkdruk onder Nederlandse huisartsen en kan bijdragen aan het toegankelijk houden van huisartsenzorg. Vandaar dat een aantal jaar geleden de APP onder de naam Extended Scope Specialist (ESS) is geïntroduceerd in Nederland. Naast verlaging van zorgkosten en verminderen van werkdruk is het behouden van kwaliteit van zorg essentieel. Een van de meest gebruikte Patient Reported Outcome Measures (PROMs) bij de beoordeling van kwaliteit van zorg is de EQ-5D. Dit instrument meet gezondheid gerelateerde kwaliteit van leven en wordt gebruikt om utiliteiten te schatten die de voorkeuren van de algemene bevolking van een land voor bepaalde gezondheidstoestanden weergeven. Deze utiliteiten zijn nodig voor de schatting van voor Quality-Adjusted Life-Years (QALY's) in kosteneffectiviteitsanalyses. Metingen van gezondheid gerelateerde kwaliteit van leven zijn echter meestal niet beschikbaar op het moment dat gegevens voor klinische doeleinden worden verzameld, zoals gegevens uit elektronische patiënt dossiers van huisartsen. Daarom verkennen onderzoekers manieren om op EQ-5D utiliteiten te voorspellen aan de hand van uitkomsten van andere gezondheid gerelateerde metingen. Dit proefschrift beoogde enkele van de uitdagingen

in de Nederlandse eerstelijnszorg te verkennen door 1) het zorggebruik en de bijbehorende kosten van huisartsenzorg bij patiënten met musculoskeletale klachten te evalueren, 2) de introductie en inzet van ESS in de eerstelijns zorg te evalueren en 3) verschillende benaderingen voor het voorspellen van ontbrekende utiliteiten te evalueren.

Deel I. Zorggebruik en kosten van musculoskeletale klachten

Hoofdstuk 2 beschrijft het zorggebruik en de kosten van door de huisarts begeleide zorg bij patiënten met klachten aan het bewegingsapparaat aan de hand van gegevens uit elektronische medische dossiers van 2.118.603 volwassen patiënten die in de huisartsenpraktijk werden behandeld. Dit aantal komt overeen met 13% van alle volwassen patiënten die bij een huisartsenpraktijk in Nederland staan ingeschreven. Daarnaast werd een General Linear Model (GLM) ontwikkeld om voorspellende factoren voor het hebben van hogere zorgkosten te identificeren. Het model werd intern gevalideerd met behulp van k-fold cross validation en de voorspellende waarde van het model werd beoordeeld met behulp van de verklaarde variantie (R^2) en root-mean-square error (RMSE). In totaal werden 403.719 patiënten geïncludeerd, waarvan 92% slechts eenmalig de huisarts consulteerde. Het aantal verwijzingen varieerde sterk voor de verschillende soorten klachten. De totale jaarlijkse zorgkosten bedroegen 39.180.531 euro, waarvan de belangrijkste kostenpost verwijzingen was, en vooral verwijzingen naar eerstelijnszorgverleners. Voor alle aandoeningen van het bewegingsapparaat samen bedroegen de gemiddelde jaarlijkse zorgkosten per patiënt €97 (SEM=€0,18). Hoge leeftijd, vrouw, lage sociaaleconomische status, wervelkolomklachten, hoog aantal musculoskeletale diagnoses en hoge comorbiditeitscore voorspelden hogere zorgkosten en verklaarden 0,7% van de variantie. Deze studie toonde aan dat de gemiddelde jaarlijkse zorgkosten van huisartsenzorg bij patiënten met aandoeningen van het bewegingsapparaat relatief laag waren en niet sterk verschilden tussen de aandoeningen. De onderzochte voorspellende factoren verklaarden een verwaarloosbaar deel van de variantie in kosten. Het is dus onduidelijk welke factoren de hoge zorgkosten in deze populatie verklaren. Toekomstige studies zouden ook kosten met betrekking tot productiviteitsverlies en informele zorg moeten bevatten.

Hoofdstuk 3 evalueert de inhoudsvaliditeit van PROMs die gebruikt worden om patiënttevredenheid te meten onder patiënten met musculoskeletale klachten binnen de eerstelijnszorg. Een systematische review werd uitgevoerd waarbij in MEDLINE, EMBASE en CINAHL (tot januari 2020) werd doorzocht om studies te identificeren over de ontwikkeling of evaluatie van de inhoudsvaliditeit van een PROM gericht op het beoordelen van patiënttevredenheid. Een PROM kwam in aanmerking als deze

gericht was op het meten van tevredenheid over de zorg bij patiënten met klachten aan het bewegingsapparaat. Twee onafhankelijke beoordelaars voerden de studieselectie, kwaliteitsbeoordeling en gegevensextractie uit. Evaluatie van de inhoudsvaliditeit van de geïncludeerde PROMs werd uitgevoerd volgens de COSMIN-richtlijnen. Er werden zeven PROMs geïdentificeerd. De kwaliteit van de ontwikkeling van alle PROMs was onvoldoende. Er werden geen validatie studies gevonden die de inhoudsvaliditeit van de geincludeerde PROMs valideren evalueerden. De inhoudsvaliditeit van de PROMs was onvoldoende en werd ondersteund door bewijs van zeer lage kwaliteit. Toekomstige studies moeten zich richten op relevantie, volledigheid en begrijpelijkheid van PROMs die gebruikt worden om tevredenheid te meten, en de nadruk leggen op betrokkenheid van patiënten tijdens de ontwikkeling van nieuwe instrumenten.

Deel II. Extended Scope Specialisten binnen de Nederlandse eerstelijnszorg

Hoofdstuk 4 beschrijft de ervaringen en percepties van ESS en huisartsen met betrekking tot de implementatie van ESS binnen de Nederlandse eerstelijnszorg. Aan deze exploratieve en interpretatieve kwalitatieve studie namen 12 ESS en 3 huisartsen deel die zich in verschillende stadia van implementatie van een ESS-zorgmodel bevonden. Tussen januari en maart 2021 werden semigestructureerde interviews afgenomen. De topiclijst was gebaseerd op bestaande literatuur, de persoonlijke inbreng van de onderzoekers en het Constellation Approach framework. De gegevens werden geanalyseerd met behulp van een thematische inductieve benadering. Uit de data kwamen vier hoofdthema's naar voren; 1) Zowel het vertrouwen van huisarts in ESS als een duidelijke toegevoegde waarde van ESS zijn cruciaal voor het starten van implementatie, 2) ESS hebben continue steun van huisartsen nodig, 3) ESS vinden dat hun positie versterkt moet worden, en 4) Implementatie van het ESS-model zorgt voor spanning over eigenaarschap. Deze vier thema's benadrukken de moeilijkheden die men ervaart bij het winnen van vertrouwen, het gebrek aan duidelijkheid over de toegevoegde waarde van ESS, de onduidelijkheid over het professionele profiel en de positionering van ESS, de behoefte van huisartsen om gezag te behouden, het gebrek aan vergoedingsstructuur, en de worsteling van ESS om een evenwicht te vinden met de huidige zorg. Deze studie toont aan dat de implementatie van een ESS-zorgmodel een uitdaging is, deels omdat de inzet van ESS niet voldoende aansluit bij de kernwaarden van huisartsen en huisartsen terughoudend lijken om de controle over patiëntenzorg over te dragen. ESS lijken geen eigenaar te zijn van de implementatie, gezien hun sterke afhankelijkheid van de praktijk, de waarden en de behoeften van de huisartsen.

Hoofdstuk 5 rapporteert de bevindingen van een exploratief onderzoek waarin de zorg voor patiënten met klachten aan het bewegingsapparaat, die traditioneel door een huisarts wordt geleverd, werd geleverd door ESS. Tussen december 2020 en december 2021 werden patiënten geïncludeerd in vier verschillende praktijken. Gegevens werden verzameld via klinische registratieformulieren en online vragenlijsten bij aanvang, 3- en 6-maanden follow-up. De kosten voor gebruikelijke zorg (i.e., zorg geleverd door de huisarts) werden berekend aan de hand van gegevens uit elektronische patiëntendossiers, waarbij gegevens werden geregistreerd door huisartsen als onderdeel van de klinische patiëntenzorg. In totaal werden 109 patiënten geïncludeerd en de meest voorkomende aandoening was schouderpijn (41%). Bij meer dan de helft van de patiënten was de duur van de klachten langer dan 6 maanden en 43% van de patiënten had een terugkerende klacht waarvoor meer dan 80% eerder een zorgverlener had geraadpleegd. Een positieve trend werd gezien voor bijna alle gezondheid gerelateerde uitkomstmaten bij 3- en 6-maands follow-up. Bijna een derde van alle door ESS uitgezette zorgpaden bevatte een verwijzing naar een tweedelijns zorginstelling. Van alle geïncludeerde patiënten werd 71% doorverwezen naar fysiotherapie, waarmee dit de meest voorkomende verwijzing was. De gemiddelde kosten van een door ESS uitgezet zorgpad bedroegen €486 (SD=€209) bij een follow-up periode van 6 maanden. De door de patiënt gerapporteerde gezondheidskosten bedroegen gemiddeld €2901 (SD=€6824) en €2729 (SD=€5715), bij een follow-up van respectievelijk 3 en 6 maanden. De gemiddelde kosten van een door de huisarts uitgezet zorgpad bedroegen €97 (SD=€117) per jaar. Een aanzienlijk deel van de patiënten die een ESS raadplegen werd gekenmerkt door langdurig terugkerende klachten en een geschiedenis van beeldvormende diagnostiek en eerdere behandeling. Bij het in kaart brengen van door ESS uitgezette zorgpaden werd een relatief hoog zorggebruik gevonden, wat verklaard kan worden door de inclusie van patiënten met complexe klachten en het stadium van ontwikkeling van de rol van ESS. De inzet van ESS laat een positieve trend zien op gezondheidsuitkomsten. Gezien het beperkte aantal deelnemende ESS en de lage aantal geïncludeerde patiënten moeten de resultaten met voorzichtigheid worden geïnterpreteerd. De bevindingen kunnen gezien worden als een startpunt voor verder onderzoek naar de inzet van ESS.

Deel III. Voorspellen van EQ-5D-3L utiliteiten

In **Hoofdstuk 6** worden verschillende regressiemodellen ontwikkeld en gevalideerd om te onderzoeken of EQ-5D-3L utiliteiten op basis van de Oswestry Disability Index (ODI) voorspelt kunnen worden en of deze voorspelde utiliteiten op een valide wijze gebruikt kunnen worden in kosteneffectiviteitsanalyses van interventies die gericht zijn op de behandeling van patiënten met lage rugpijn. Individuele patiëntgegevens van 18.692 patiënten met lage rugpijn werden gesplitst in een trainings- en een validatie

groep. In totaal werden zes verschillende modellen ontwikkeld: 1) Ordinary Least Squares (OLS) regressie, met totale ODI score, 2) OLS, met ODI item scores als continue variabelen, 3) OLS, met ODI item scores als ordinale variabelen, 4) Tobit model, met totale ODI score, 5) Tobit model, met ODI item scores als continue variabelen, 6) Tobit model, met ODI item scores als ordinale variabelen. EQ-5D-3L utiliteiten van werden opgenomen als onafhankelijke variabele en ODI-scores als onafhankelijke variabele. Het voorspellend vermogen van de modellen werden beoordeeld aan de hand van de verklaarde variantie (R^2) en de Root Mean Squared Error (RMSE). De mogelijke impact van het gebruiken van voorspelde, in plaats van werkelijk gemeten, EQ-5D utiliteiten op de kosteneffectiviteitsuitkomsten werd geëvalueerd in twee empirische kosteneffectiviteitsanalyses.

Alle modellen hadden een min of meer vergelijkbare R^2(range: 45-52%) en RMSE (range: 0,21-0,22). De twee best presterende modellen (i.e. model 2 en 5) lieten relatief vergelijkbare voorspelde kansen op kosteneffectiviteit zien bij verschillende referentiewaarden vergeleken met de geobserveerde waarden. Het verschil in kansen varieerde bijvoorbeeld van 2% tot 5% bij een referentiewaarde van 50.000 €/QALY. De resultaten suggereren dat de ODI op een valide wijze kan worden gebruikt om de EQ-5D-3L utiliteiten en QALY's van lage rugpijn patiënten te voorspellen voor gebruik in kosteneffectiviteitsanalyses op het moment dat de EQ-5D niet is afgenomen. Aanvullende onderzoek is nodig om 1) de modellen te valideren om na te gaan of deze modellen vergelijkbare resultaten opleveren in andere empirische datasets over lage rugpijn interventies, 2) de generaliseerbaarheid van de geschatte modellen te verbeteren, en 3) de prestaties van voorspellende modellen te vergelijken met een response mapping modellen voor het schatten van utiliteiten. Ondertussen kunnen onderzoekers de ontwikkelde modellen gebruiken in hun kosteneffectiviteitsanalyse wanneer utiliteiten ontbreken.

Hoofdstuk 7 heeft als doel het ontwikkelen en valideren van verschillende response mapping modellen om EQ-5D-3L utiliteiten te voorspellen op basis van Oswestry Disability Index (ODI) scores en om deze modellen vervolgens te vergelijken met de regressie modellen die ontwikkeld zijn in *Hoofstuk 6*. In deze studie werd dezelfde data gebruikt als tijdens het ontwikkelen en valideren van de eerder besproken regressie modellen. Drie verschillende respons mapping modellen werden ontwikkeld: 1) non-parametrisch model (Non-p), 2) non- parametrisch model, exclusief logische inconsistenties (non-peLI) en een 3) ordinale logistische regressie model (ORL). Het voorspellend vermogen van de modellen werden beoordeeld aan de hand van de verklaarde variantie (R^2), de Root Mean Squared Error (RMSE) en Mean Absolute Error (MAE). Om te evalueren of de MAE's en de bijbehorende limits of agreement (95%LA) klinisch relevant waren,

werd een minimally clinically important difference (MCID) van 0,074 aangehouden. De Non-p bleek het beste voorspellende vermogen te hebben (R2=0,43; RMSE=0,22; MAE=0,03; 95%LA=-0,40;0,47) vergeleken met de NonpeLI (R2=0,07; RMSE=0,29; MAE=-0,15; 95%LA=-0,63;0,34), en de ORL (R2=0,22; RMSE=0,26; MAE=0,02; 95%LA=-0,49;0,53). MAE's waren kleiner dan de MCID voor de Non-p en de OLR, maar niet voor de Non-peLI. De verschillen in de kans op kosteneffectiviteit liepen uiteen van 1% tot 4% (Non-p), 0,1% tot 9% (Non-peLI) en 0,1% tot 20% (OLR). De resultaten suggereren dat de ontwikkelde response mapping modellen niet valide zijn om EQ-5D-3L utiliteiten van individuele patiënten te schatten. Daarnaast kan het gebruik van sommige mapping modellen een significante invloed kan hebben op de resultaten van een economische evaluatie. Verder bleek het voorspellend vermogen van de response mapping modellen niet beter te zijn dan die van de regressiemodellen welke ontwikkeld zijn in *Hoofdstuk 6*. Dit maakt dat de ontwikkelde response mapping modellen niet worden aanbevolen voor gebruik in economische evaluaties.

Discussie

In **Hoofdstuk 8** worden de belangrijkste bevindingen van dit proefschrift besproken en geïnterpreteerd op basis van bestaande literatuur. De meeste patiënten met een musculoskeletale aandoening brachten eenmalig een bezoek aan de huisarts en de gemiddelde jaarlijkse zorgkosten van huisartsenzorg waren relatief laag. Verwijspercentages varieerden tussen de verschillende klachten en lijken in overeenstemming met aanbevelingen uit de richtlijnen van het Nederland Huisartsen Genootschap. Dit suggereert dat huisartsen zich grotendeels aan deze richtlijnen houden en benadrukt het belang van richtlijnontwikkeling en implementatie. Een beperkt aantal beschikbare voorspellende factoren in combinatie met een relatief lage variatie in zowel de voorspellende factoren als de uitkomstvariabele (i.e., zorgkosten) leidde tot een laag voorspellend vermogen van het regressiemodel met statistisch significante, maar relatief kleine bèta's. In een grote populatie kan de impact op het zorgbudget echter groot zijn ondanks de kleine bèta's. Onderzoekers moeten zich bewust zijn van validiteitsproblemen en beperkingen van het gebruik van gegevens die niet voor onderzoeksdoeleinden zijn verzameld, zoals klinische registratiegegevens. De implementatie van een ESS-zorgmodel in de Nederlandse eerstelijnszorg is een uitdaging gebleken, omdat ESS geen eigenaarschap lijken te hebben over de implementatie, gezien hun sterke afhankelijkheid van de praktijkvoering, waarden en behoeften van huisartsen. De implementatie van een ESS-zorgmodel in de huisartsenpraktijk zou gebaat zijn bij steun van huisartsen en andere belanghebbenden, een duidelijk omschreven rol, standaardisatie van processen en werkmethoden, beschikbaarheid van training op een passend niveau, afstemming op

de kernwaarden van huisartsen en een duidelijke meerwaarde van APP ten opzichte van al bestaande samenwerkingsverbanden tussen huisartsen en fysiotherapeuten. Een aanzienlijk deel van de patiënten die een ESS raadpleegden werd gekenmerkt door langdurig terugkerende klachten en een geschiedenis van diagnostische beeldvorming en eerdere behandeling. Bij het in kaart brengen van door ESS uitgezette zorgpaden werd een relatief hoog zorggebruik gevonden, wat verklaard kan worden door de inclusie van patiënten met complexe klachten en het stadium van ontwikkeling van de rol van ESS. Ten slotte, wanneer EQ-5D-3L utiliteiten ontbreken tijdens het uitvoeren van een economische evaluatie kunnen ODI-scores gebruikt worden om deze missende utiliteiten te schatten. Hierbij laten regressie modellen een beter voorspellend vermogen zien dan response mapping modellen. De ontwikkelde modellen zijn niet geschikt voor het voorspellen van utiliteiten van individuele patiënten. Ondanks dat EQ-5D-3L utiliteiten met behulp van de ODI geschat kunnen worden, is het afnemen van EQ-5D nog steeds de gepaste werkwijze binnen economische evaluaties.


11

Acknowledgements | Dankwoord

Dankwoord | Acknowledgements

Vier jaar gelden startte ik mijn promotietraject. Ondertussen zijn de jaren zijn voorbijgevlogen en zit het erop. In al die jaren heb ik het ontzettend leuk gehad, ondanks dat het niet altijd even makkelijk was om buiten mijn comfort zone te werken (en ik een beetje routinewerk op zijn tijd ben gaan waarderen ☺). Gelukkig heb ik dit alles niet alleen hoeven doen. Ik heb van vele getalenteerde collega's mogen leren en ben al die tijd door lieve mensen in mijn omgeving gesteund en aangemoedigd. Een 'big thank you' is dan ook op zijn plaats!

Als eerste wil ik graag mijn promotieteam bedanken. **Professor Pool-Goudzwaard, professor Ostelo** en **professor van Tulder,** ik ben heel dankbaar dat ik bij jullie mocht promoveren. De ruimte en het vertrouwen dat jullie mij de afgelopen jaren hebben gegeven waardeer ik enorm. Mijn complimenten dat jullie het al die jaren met mijn ongeduld en eigenwijze trekken hebben volgehouden! Beste **Annelies**, ik denk dat we het er allemaal wel over eens zijn dat ESS niet zover gekomen zou zijn zonder jouw jarenlange en tomeloze inzet. Dank dat je mij op sleeptouw hebt genomen, en naast al jouw inhoudelijke adviezen er ook voor zorgde dat we veel met elkaar hebben gelachen ☺. Beste **Maurits**, jij bent al die jaren zo'n betrouwbare en constante factor geweest. Dit heeft mij veel rust gegeven op momenten dat het minder makkelijk liep. En niet te vergeten, jij en ik waren altijd als eerste op Zoom! Beste **Raymond**, wat fijn dat je altijd kritisch mijn stukken leest en om verduidelijking vraagt. Jij hebt mij geleerd dat om niets zomaar aan te nemen en scherp te blijven, een eigenschap die altijd goed van pas komt ☺.

Beste **dr. Scholten-Peeters**, **prof. van der Wees**, **prof. Veenhof**, **prof. Schers** en **dr. van Nassau**, hartelijk dank dat jullie bereid waren om mijn proefschrift te beoordelen en mij te voorzien van waardevolle suggesties.

Beste **Marianne**, dank voor al jouw hulp tijdens het uitvoeren kwalitatieve studie. Zonder jou had ik het moeras van open codes niet overleefd. Dank dat je mij bewuster hebt gemaakt van het belang van het proces in plaats van het resultaat. De weg is beter dan de herberg.

Beste **Lamyae**, wat fijn om met jou te mogen samenwerken. Jouw frisse blik heeft ons artikel zo veel beter gemaakt! Promoveren, ook wat voor jou?!

Beste **Eddy**, dank dat je mij hebt geïntroduceerd in het domein van huisartsgeneeskunde. Het heeft mij niet alleen inhoudelijk meer inzicht en verdieping gegeven, maar ook laten zien dat huisartsgeneeskunde een ontzettend mooi vak is.

Beste **Hanneke**, hoe kan een promovendus zonder jou?! Bij jou kon ik altijd terecht. Heel knap hoe jij moeilijke dingen makkelijk maakt en allemaal op zo'n relaxte manier ☺. Onwijs bedankt voor alles!!!

Dear **Angela**, I feel very privileged working with you. You helped me to get started with R and later with debugging complicated codes. I honestly could not have done this without you. You are the most dedicated, modest but foremost lovely person!

Beste **Judith**, ondanks dat ik niet echt bij jouw onderzoeksgroep behoor, heb ik me altijd heel welkom gevoeld bij jullie. Dank hiervoor!!

Lieve **Juul**!!! 1000 maal dank voor alle leuke momenten en jouw lekkere nuchtere blik. Dat had ik never niet willen missen. Nog even en ik zit bij jouw verdediging ☺.

Beste **Esther,** niet normaal hoe hard en toegewijd jij werkt. Jij kan bergen verzetten! Met jou (jullie ;-)) het WHO-project doen vond ik echt superleuk!!

Beste **Sidney**, dank voor de fijne samenwerking tijdens het WHO-project. Je mag echt supertrots zijn op alle drie de reviews die we met elkaar hebben afgeleverd!

Beste **Annemarie**, dank voor al jouw hulp tijdens het WHO-project. Ik denk dat ik niet eens de helft heb meegekregen van wat jij allemaal aan werk gedaan hebt!

Dear **Tiziano**, thank you for all your hard work on the WHO project. Good luck with wrapping up your PhD. Hoping for a 'defense-after-party' in Italy, so we can finally get some good weather an Italian food ;-)

Dear **Alessandro**, my first paper was with you. Thank you for introducing me into the world of clinimetrics and getting me up to speed with my promotion!

Beste leden van het **NVES-bestuur**, hartelijk dank voor de fijne samenwerking. Ik wens jullie veel succes in de toekomst. Ik ben benieuwd hoe en waar ESS zijn plek gaat vinden.

Dank aan alle bevlogen ESS en huisartsen in het veld die ik de afgelopen jaren gesproken heb, maar in het bijzonder de ESS (en huisartsen) die hebben deelgenomen aan de observationele studie: **Marco**, **Karin**, **Patrick**, **Marijn** en **Rikie** heel erg veel dank voor jullie hulp! Alle hordes die we hebben moeten nemen en de lange aanloop naar de start van het onderzoek. Zonder jullie was dit niet mogelijk geweest!!

Dank aan alle **deelnemers van de kwalitatieve studie** (helaas kan ik jullie niet bij naam kan noemen) voor ontzettende leuke en leerzame interviews. Dit heeft veel nieuwe en goede inzichten gebracht!

Lieve **Lin**, mijn allerbeste vriendin, of zoals de kinderen jou noemen, jouw zus Linda en dat ben jij eigenlijk ook, aangezien wij al bij elkaar zijn zo lang als ik me kan herinneren en dat is ondertussen eigenlijk best wel lang ;-) Dank voor al jouw steun in de 4(0) jaar. Diep respect heb ik voor jouw onuitputtelijke en onvermoeibare inzet op het thuisfront. Daar is 4 jaartjes PhD-en niks bij ☺

Lieve **Luitzen**, de master epidemiologie heeft veel gebracht, maar het allermooiste dat ik eraan over heb gehouden is onze vriendschap. Nu dat mijn proefschrift is afgerond is het jouw beurt!!!

Beste **Olaf**, mede dankzij jou ben ik begonnen aan mijn master epidemiologie. De voorloper van mijn promotie. Dank dat je mij geleerd hebt om naar de grote lijnen te kijken. Details, Tom Poes!

Lieve **pa**, jij en ma hebben mij geleerd wat hard werken is en altijd gestimuleerd om naar school te gaan en iets te leren zo lang als het kan. Waar ik nu ben en wat ik doe is dankzij jullie. Lieve **ma**, ook al ben je niet meer bij als ik mijn proefschrift mag verdedigen, ik weet zeker dat je meekijkt en trots op mij bent.

Mijn allerliefste kinderen, wat ben ik blij met jullie!! **Link**, ik bewonder de manier waarop jij dicht bij jezelf blijft en je niet laat leiden door wat andere mensen vinden. **Fox**, de manier waarop jij andere mensen samenbrengt en gezien kan laten voelen is echt een hele bijzondere eigenschap, daar mag je trots op zijn. **James**, superknap hoe jij eindeloos weet door te zetten met een grote lach op je gezicht, ook als het moeilijk gaat! Van alle mensen om mij heen heb ik van jullie het meest geleerd. Ik hou van jullie!! Knuffel kan altijd ☺

Lieve **Alwin**, dank dat je mij alle ruimte geeft om mijn eigen ding te doen, maar mij ook laat inzien dat ik niet per se alles alleen hoef te doen. Ik ben blij dat jij het al die jaren met mij volhoudt, ondanks mijn innerlijke onrust en eigenwijze karakter (never a dull moment ☺). Luv you en nu.... op naar de boot!!


12

List of publications

Publications for this thesis

Pellekooren S, Ostelo RWJG, Pool-Goudzwaard AL, van Tulder MW, Jansma E, Chiarotto A. Content Validity of Patient-Reported Outcome Measures of Satisfaction with Primary Care for Musculoskeletal Complaints: A Systematic Review. J Orthop Sports Phys Ther. 2021. doi:10.2519/jospt.2021.9788

Pellekooren S, Donker M, Reijnders E, Boutalab L, Ostelo RWJG, van Tulder MW, Pool-Goudzwaard LA. The introduction of advanced practice physiotherapy within Dutch primary care is a quest for possibilities, added value, and mutual trust: a qualitative study amongst advanced practice physiotherapists and general practitioners. BMC Health Serv Res. 2022. doi:10.1186/s12913-022-07906-6

Pellekooren S, Ben ÂJ, Bosmans JE, Ostelo RWJG, van Tulder MW, Maas ET, Huygen FJPM, Oosterhuis T, Apeldoorn AT, van Hooff ML, van Dongen JM. Can EQ-5D-3L utility values of low back pain patients be validly predicted by the Oswestry Disability Index for use in cost-effectiveness analyses? Qual Life Res. 2022. doi:10.1007/s11136-022-03082-6

Pellekooren S, Ostelo RWJG, van Tulder MW, Pool-Goudzwaard AL. Implementation of an Advanced Practice Physiotherapy model of care in Dutch primary care; an explorative observational study. 2022. (Report for the Dutch Advanced Physiotherapy Practitioners Association; Nederlandse Vereniging voor Extended Scope Specialisten) (In Dutch).

Ben ÂJ, Pellekooren S, Bosmans JE, Ostelo RWJG, Maas ET, El Alili M, van Tulder MW, Huygen FJPM, Oosterhuis T, Apeldoorn AT, van Hooff ML, van Dongen JM. Mapping Oswestry Disability Index responses to EQ-5D-3L utility values: are cost-utility results valid? Value in Health.2023. doi: 10.1016/j.jval.2023.01.020

Pellekooren S, Ben ÂJ, van Dongen JM, Pool-Goudzwaard AL, van Tulder MW, van den Berg JM, Ostelo RWJG. Predicting Direct Healthcare Costs of General Practitioner Guided Care in Patients with Musculoskeletal Complaints. Pain, 2023. doi:10.1097/j.pain.0000000000003028

Other publications

Ratter J, Pellekooren S, Wiertsema S, van Dongen JM, Geleijn E, de Groot V, Bloemers FW, Jansma E, Ostelo RWJG. Content validity and measurement properties of the Lower Extremity Functional Scale in patients with fractures of the lower extremities: a systematic review. J Patient Rep Outcomes. 2022. doi:10.1186/s41687-022-00417-2

Pellekooren S, Maas ET, Ostelo RWJG, Innocenti T, Chiarotto A, Annemarie de Zoete A, Rubinstein SM The effect of traction in people with chronic primary low back pain: a systematic review. WHO Guidelines on management of chronic primary low back pain in adults. 2022.

Maas ET, Pellekooren S, Ostelo RWJG, Innocenti T, Chiarotto A, de Zoete A, Rubinstein SM. The effect of massage in people with chronic primary low back pain: a systematic review. WHO Guidelines on management of chronic primary low back pain in adults. 2022.

De Zoete A, Innocenti T, Maas ET, Pellekooren S, Chiarotto A, Ostelo RWJG, Rubinstein SM. The effect of spinal manipulative therapy in people with chronic primary low back pain: a systematic review. WHO Guidelines on management of chronic primary low back pain in adults. 2022.


13

PhD Portfolio

PhD Portfolio

For the PhD thesis preparation, the courses below were followed as part of the Training Program of the Amsterdam Movement Sciences Research Institute.

	Date Achieved	EC
Courses		
Research ethics (Vrije Universiteit Amsterdam)	April 2020	3
Basiscursus Regelgeving & Organisatie voor Klinisch Onderzoekers (NFU)	Summer 2019	1.5
Writing a Data Management Plan (Vrije Universiteit Amsterdam)	December 2020	1
Advanced academic writing for PhD researchers (Vrije Universiteit Amsterdam)	June 2019	3
PhD success and personal efficacy (Vrije Universiteit Amsterdam)	October 2019	2
University Teaching Program (BKO, Vrije Universiteit Amsterdam)	January 2020	5
Economic Evaluations (Vrije Universiteit Amsterdam AM_470828)	October 2020	6
Basic concepts in qualitive research in health care (University of Amsterdam)	June 2019	5
Introduction to Bayesian Statistics (Amsterdam UMC EpidM R84)	April 2022	2
Research related		
Amsterdam Movement Sciences annual meeting (poster presentation)	March 2020	2
Amsterdam Movement Sciences annual meeting (poster presentation)	March 2022	2
Jury member Amsterdam Movement Sciences Research Institute poster presentations	March 2021	1
Back and Neck Pain Forum (poster and oral presentation)	November 2021	4
Total number of ECTS		**41.5**


14

About the author

About the author

Sylvia Pellekooren was born in 1981 in Boskoop. In 2000, she started studying physiotherapy in Leiden and graduated in 2004. She started working as a clinical physiotherapist in various hospitals for a short period of time. In 2006, she started her specialisation as a hand therapist at Erasmus Medical Centre and shortly thereafter became a certified hand therapist (CHT-NL). Besides working with patients, she was active in teaching, working on anatomic models, and research. Having enjoyed working as a hand therapist for many years she started the master Evidence Based Practice in Health care in 2016. After graduating in 2018 she started her PhD trajectory in 2019 at the faculty of Human Movement Sciences of the Vrije Universiteit Amsterdam.

In 2023, she started working at department of Efficiency Studies of ZonMw in The Hague.