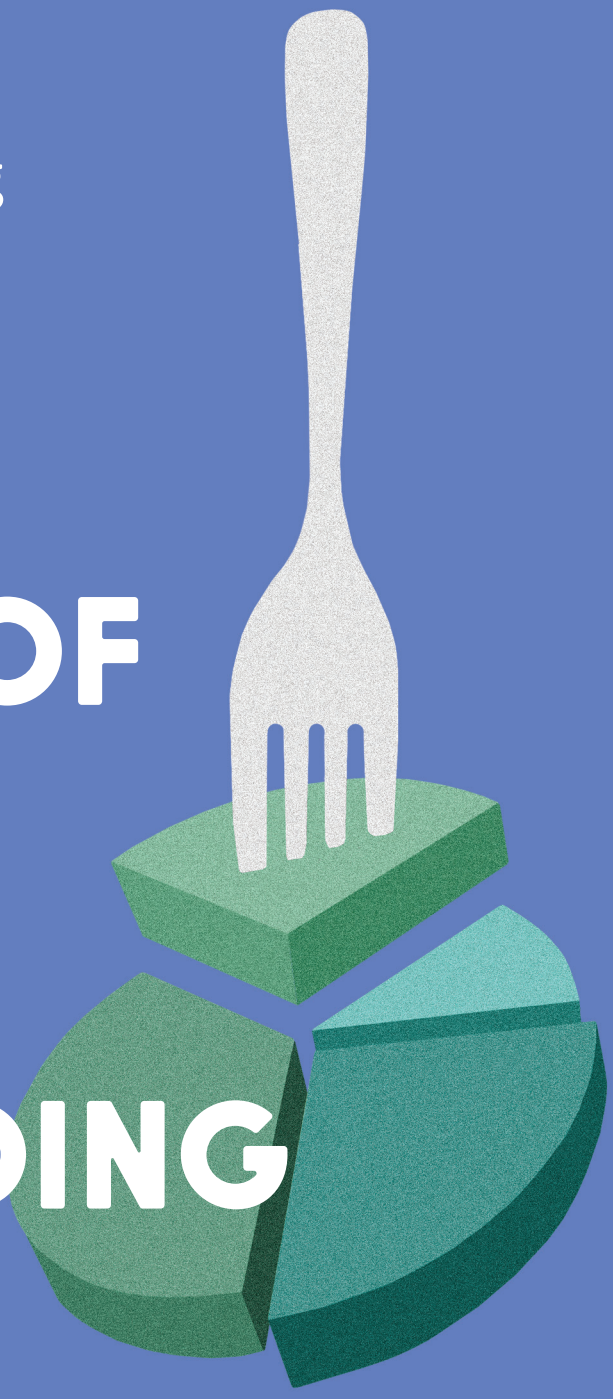
Koen Verburg

THE PROOF OF THE PUDDING



Using outcome-based quality indicators in physical therapy

The proof of the pudding:
Using outcome-based quality indicators in physical therapy

Koen Verburg

Cover illustration	Martien ter Veen
Design	Anna Bleeker | persoonlijkproefschrift.nl
Print	Ipskamp Printing | proefschriften.net
ISBN	978-94-6421-795-7

The work presented in this thesis was carried out within the Department of the Scientific Centre for Quality of Healthcare (IQ healthcare) part of the Radboud Institute for Health Sciences of the Radboud university of medical center, Nijmegen, the Netherlands

The printing of this theses was financially supported by the Radboud Institute for Health Sciences of the Radboud university medical center, Dutch Patient Association for Back Pain (NVVR), the Lung Foundation Netherlands (Long Fonds), the Association for Quality in Physiotherapy (SKF), the Royal Dutch Society for Physiotherapy (KNGF), Geeresteingroep, health insurance companies CZ & De Friesland

The proof of the pudding:
Using outcome-based quality indicators in physical therapy

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

vrijdag 9 september 2022
om 12:30 uur precies

door

Arie Cornelis Verburg
geboren op 7 april 1991
te Amersfoort

Promotoren

Prof. dr. P.J. (Philip) van der Wees

Prof. dr. M.W.G. (Ria) Nijhuis-van der Sanden

Copromotoren

dr. S.A. (Simone) van Dulmen

dr. H. (Henri) Kiers, Hogeschool Utrecht

Manuscriptcommissie

Prof. dr. H.J. (Henk) Schers

Prof. dr. M. (Mark) van Houdenhoven

Prof. dr. R.A. (Rob) de Bie, Maastricht University

Paranimfen

Laura van Kempen

Job Klumper

The proof of the pudding:
Using outcome-based quality indicators in physical therapy

Dissertation to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. J.H.J.M. van Krieken,
according to the decision of the Doctorate Board
to be defended in public on

Friday, September 9, 2022
at 12.30 pm

by

Arie Cornelis Verburg
born on April 7, 1991
in Amersfoort (the Netherlands)

Supervisors

Prof. dr. P.J. (Philip) van der Wees

Prof. dr. M.W.G. (Ria) Nijhuis-van der Sanden

Co-supervisors

dr. S.A. (Simone) van Dulmen

dr. H. (Henri) Kiers, HU University of Applied Sciences Utrecht

Manuscript Committee

Prof. dr. H.J. (Henk) Schers

Prof. dr. M. (Mark) van Houdenhoven

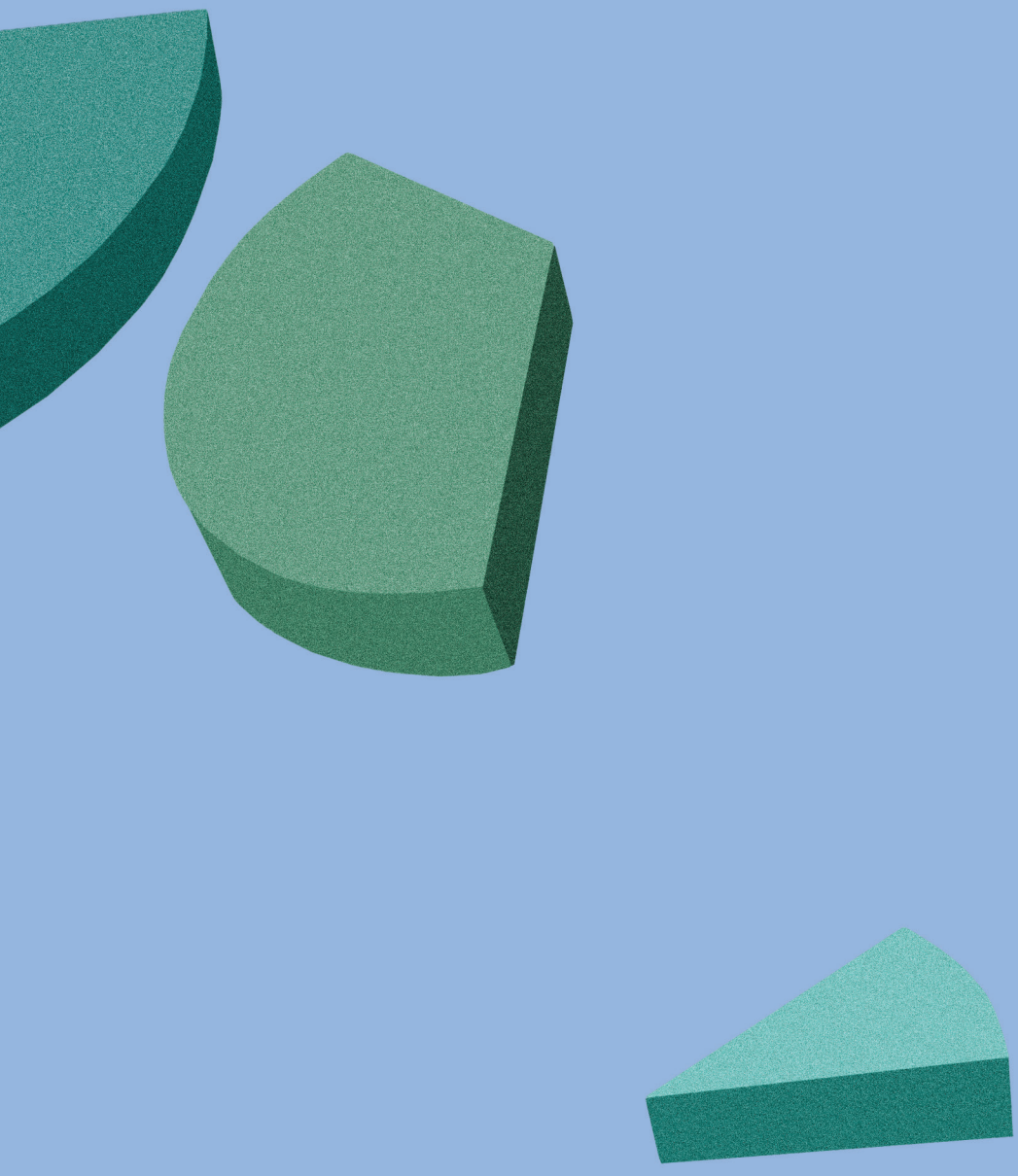
Prof. dr. R.A. (Rob) de Bie, Maastricht University

Paranymphs

Laura van Kempen

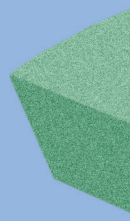
Job Klumper

CONTENTS



CHAPTER 1

General introduction



BACKGROUND

Quality of care has been defined as: *"doing the right thing, at the right time, in the right way, for the right person, and having the best possible results"*.(1) The challenge is to determine how to quantify and gain insights into quality of care in daily practice. Over the past few decades, the monitoring of quality of care was mainly focussed on evaluating processes of care, i.e., doing the right thing, at the right time, in the right way.(2-7) Examples of processes of care include measuring guideline adherence, registering the diagnostic process, or administrating intervention goals. In recent years, health care has begun to shift towards estimating the value of treatments for individual patients, i.e., focussing on the right person for the best possible results.(7, 8)

Recording the meaningful outcomes of patients provides several opportunities to monitor, evaluate and improve quality of care (see figure 1). First, outcome measurements can be used to encourage interaction between patients and health providers to improve clinical care, e.g., for goal setting purposes and shared decision-making. Second, by aggregating meaningful outcomes at the group level, healthcare providers can reflect on their own performance and compare health outcomes with peers. Third, outcomes at the group level can also be used as reference values to compare individual outcomes. Fourth, the aggregation of outcomes can also be used for public reporting, e.g., as a tool for helping patients to choose a provider or for pay-for-performance initiatives.

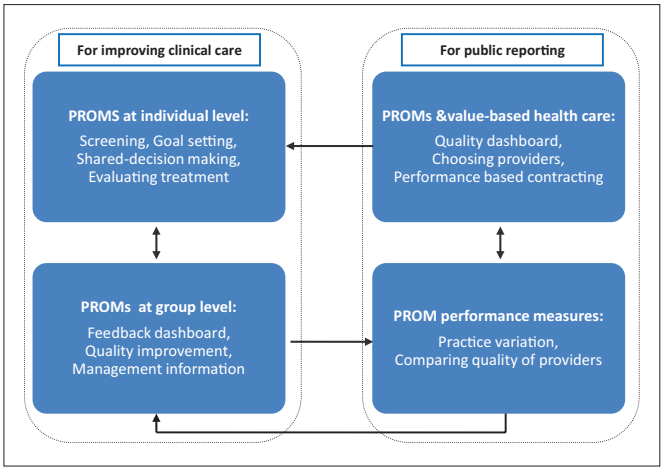


Figure 1: Framework for the innovative multipurpose use of patient-reported outcome measures (PROMs), described by van der Wees (9)

In daily practice, the collection of real-world data poses several challenges for quality-improvement initiatives. In 2014, experts from three countries stated that,

although the key conclusion was that using patient outcomes for quality improvement is feasible, "providers, patients, and purchasers of care must agree on a common vision. Building trust among stakeholders that the data will not be misused seems fundamental to success".(10) Stakeholder engagement in developing quality-improvement initiatives is therefore an important condition for their successful implementation.(11-14) When standardised provided data are compared between healthcare providers and are perceived as useful in daily practice, opportunities arise for informed decision-making and other quality-improvement initiatives.(14, 15)

Furthermore, the routine collection of sufficient, valid and reliable data by many healthcare providers requires a substantial financial, organisational and implementational investment.(10, 16, 17) Standardisation and the use of core sets of meaningful outcomes derived from electronic health records (EHRs) keeps the investment for data collection as low as possible; however, it is vital that healthcare providers and patients experience the benefits of routine data collection in daily practice.(10, 17)

This thesis aims to develop, select and test a core set of outcome-based quality indicators in primary care physical therapy practice for patients with non-specific low back pain (NSLBP) and chronic obstructive pulmonary disease (COPD). As an introduction to the research aims of this thesis, we first outline current (inter) national outcome-based quality-improvement initiatives, specifically those for primary physical therapy care. Furthermore, we will introduce the patient-reported outcome measures (PROMs), quality indicators and patient populations included in our research. The introduction will end with the research questions and thesis outline.

International and national initiatives focussed on health outcomes

In 2006, Michael E. Porter introduced the Value-Based Health Care (VBHC) model as a blueprint for reforming health care.(18) The goal of the VBHC is to achieve high-value care for patients, with value defined as the health outcomes relative to the costs for achieving these outcomes.(8, 18) Porter stated that "…this goal is what matters to patients and unites the interests of all actors in the system. If value improves, patients, payers, providers and suppliers can all benefit while the economic sustainability of the healthcare system increases".(8) Since value depends on the results and not merely on the delivery of care, value in health care is measured by the outcomes achieved, while the volume of services delivered are part of the costs. This shift in focus from volume to value is a central challenge;(8) for example, how can the comparability of the outcomes between patients or providers be ensured in order to measure value?

Such challenges can be reduced by standardising outcomes, as is the aim of the International Consortium for Health Outcomes Measurement (ICHOM).(15) ICHOM is an international organisation that creates standardised outcome sets, based on

consensus procedures with stakeholders, for a valid comparison of outcomes between patients or providers.(15) ICHOM standard outcome sets also take into account the importance of risk adjustment by including case-mix variables. The outcome sets can be used for shared decision-making and goal setting in individual patient care, and for improving performance or value-based payment purposes. To date, ICHOM has published more than 30 standard sets for different patient populations.(15)

In the Netherlands, various initiatives concerning the collection and aggregation of health outcomes are being conducted. In 2018, the Ministry of Health, Welfare and Sports (VWS) started a national programme focussed on measuring outcomes of care to improve the quality of life of patients, and to increase the quality of healthcare providers.(19) VWS launched four main topics: 1) improving transparency in outcomes, 2) improving shared decision-making, 3) focussing on outcome in the organisation and cost of health care and 4) better access to relevant and actual outcome information.

An example of a VBHC initiative in the Netherlands is a project of the Santeon network, which is called “open benchmark”. Santeon is a network of seven large hospitals that aims to monitor, compare and potentially improve their quality of care, focussing on patient centredness and collecting meaningful outcomes for patients.(20) Using open benchmark, several processes and outcomes are aggregated for each department and compared in order to learn from the differences and improve care with the help of dashboards and scorecards. They compare processes such as the number of full-time equivalent members of staff (FTEs) and outcomes such as differences in the value of delivered care. Santeon is currently conducting several projects with different patient populations using quality-improvement cycles.(20)

Initiatives in Dutch physical therapy

Dutch primary care physical therapists are also using outcomes for quality improvement, such as through the use of national data registries. These national data registries are co-ordinated by the professional bodies in Dutch physical therapy, the Royal Dutch Society for Physical Therapy (KNGF) and the Association for Quality in Physical Therapy (SKF).(21, 22) In 2013, a large programme, “Quality in Motion”, was launched by the KNGF, with the aim of improving patient centeredness and the effectiveness of care through the use of patient health outcomes.(21) From there, several pilots were launched, collecting data on the structure, process and outcomes of care via the EHRs derived from routinely collected real-world data,(21) which can be used in the development of quality indicators. Quality indicators are measurable items within the structure, processes or outcomes of care that can be used to monitor health care.(23) Quality indicators based on health outcomes can be used as a tool to improve quality of care by monitoring and evaluating treatment trajectories.(14, 21, 24, 25)

The national data registries and initiatives such as Quality in Motion are important for comparing outcomes between physical therapists; however, questions remain surrounding the estimation of the value of care based on routine outcome measurements: which outcome domains and measures are relevant and feasible to monitor and improve physical therapy treatment? What is the comparability of the treatment outcomes between physical therapists or physical therapy practices based on real-world observational data? What are the views of the end-users (patients, physical therapists) regarding the usefulness of the quality indicators and do they accept them as a quality-improvement tool? How should the reliability of the data collection be increased?

PROMs and physical-performance measures

Outcomes of care can be collected as PROMs, which are questionnaires or single-item scales used to assess patient-reported outcomes in relevant outcome domains, such as pain, physical functioning, or perceived treatment effect.(23, 26) For example, the Quebec Back Pain Disability Scale (QBPDS) is a 20-item questionnaire used to measure physical functioning and limitations in patients with low back pain.(27) An important requirement for using outcomes for quality improvement is the selection of outcome domains and associated PROMs that are meaningful for the patient population treated in primary care physical therapy practice. PROMs are often combined with other clinician-assessed, impairment-based or physical-performance measures to provide a more complete interpretation of patient outcomes and impairments.(28) Examples of physical performance measures are the six-minute walk test (6MWT) for measuring physical capacity, and the hand-held dynamometer for measuring quadriceps strength.

Quality indicators

As described earlier, quality indicators are measurable items used to monitor the structure, processes or outcomes of health care.(23) The selection of meaningful outcomes for the development of quality indicators should be based on evidence-based healthcare recommendations, e.g. derived from clinical practice guidelines (CPGs).(29-32) After developing the quality indicators, consensus meetings are usually conducted between stakeholders to select a final set of quality indicators.(29-31)

Depending on the purpose of the quality indicators, the perspectives of the stakeholders may differ. Eligible stakeholders in physical therapy care are patients, physical therapists, practice managers, professional associations, health insurance companies, policy-makers and politicians.(33) It is therefore highly important that a structured consensus procedure is used to select quality indicators perceived to have added value for every stakeholder.

Primary care settings in the Netherlands

In the Netherlands, primary care has a leading role in health care for the Dutch population. An aspect of the Dutch primary care system is that patients can access physical therapy services directly or following a referral by a general practitioner.(34) In 2006, Dutch health care was restructured to universal coverage through a mandatory basic health insurance package and private additional insurance, merging public and private insurance. Patients purchase their basic and additional coverage from private insurance companies.(34, 35)

Primary care physical therapists receive their reimbursement through a fee-for-service system,(34) in which health insurers play a key role in monitoring and stimulating quality of care by setting conditions for contracting physical therapy practices that provide meaningful treatment for their insured patients. One example is monitoring the treatment frequency with the so-called 'treatment index', which provides insight into the mean treatment frequency of a physical therapy practice in comparison with that of all physical therapy practices adjusted for some case-mix patient characteristics.(36) In the Netherlands, the treatment index is standardised by all health insurers.(36) Physical therapists shared mixed feelings about whether the treatment index is a useful tool for the evaluation of quality of care however,(37) because the amount of treatment sessions gives no insight into the experienced effects of the treatment or the recurrences of patients. Many stakeholders are therefore searching for other methods that enable a more complete interpretation of the quality of delivered care.

Study population in this thesis

Here, we focus on developing, selecting and testing core sets of outcome-based quality indicators based on the outcome domains and associated measures for patients with NSLBP and COPD in primary care physical therapy, a common type of care for these patient groups. Both patient populations are heterogenous, with large variations in their experienced health status.(38-40)

At both the national and international level, several studies have reported on the development of quality indicators in physical therapy care.(4-6, 24, 28, 29, 41-44) These studies did not focus on primary physical therapy care for patients with NSLBP or COPD however, but rather patients with neck pain, peripheral artery occlusive disease, Parkinson's disease, and those undergoing hip and knee arthroplasty rehabilitation for hip and knee osteoarthritis and rheumatoid arthritis. Furthermore, these studies mostly focussed on the evaluation of processes of care, while health care is shifting towards the estimation of the value of treatments for individual patients by monitoring treatment outcomes. It remains unknown which outcomes for patients with NSLBP or COPD could be used for the development of quality indicators and for quality-improvement purposes, as described in figure 1.

Non-specific low back pain (NSLBP)

According to the *Lancet* series, published in 2018, low back pain is now the leading cause of disability worldwide.(45) In most cases, it is not possible to identify a specific nociceptive cause, in which case the low back pain is considered nonspecific (NSLBP).(45) In this thesis, NSLBP is further defined as pain and discomfort localised below the costal margin and above the inferior gluteal folds, with or without leg pain, and not caused by a specific pathology.(46) Patients with NSLBP experience a high variation in the level of recovery and time to recover, varying from one day to multiple years. (46)

In the Netherlands, patients with NSLBP are among the most treated patients in primary care physical therapy practice.(47) Many treatment options exist for patients with NSLBP.(48) Currently, only limited data is available for standardised treatment outcomes and the quality of daily care of patients with NSLBP treated in physical therapy practices in the Netherlands.(49) In most cases, this research is conducted in a controlled environment, such as randomised controlled trials, or in studies focussed on evaluating the processes of care rather than the outcomes.(50-53)

In 2013, the Dutch clinical practice guideline for patients with low back pain was published, which includes recommendations for meaningful outcome domains and associated measures.(54) This could be used as a basis for the selection of outcome domains and measures for the further development of quality indicators, supported by other (inter)national literature.(25, 32)

Chronic obstructive pulmonary disease (COPD)

According to the World Health Organization (WHO), COPD is the fourth most common cause of death worldwide, and is predicted to become the third most common by 2030.(55) COPD is an incurable, progressive but manageable respiratory disease. (56, 57) Patients with COPD often report having a low physical capacity and/or a low physical activity, which are important areas for improvement through primary care physical therapy treatment.(58-61)

For patients with COPD, little is known about the outcomes of Dutch daily primary physical therapy practice; however, research has shown that physical therapy rehabilitation interventions can be beneficial for these patients, particularly in terms of improving muscle function and cardiovascular function and reducing exacerbations. (59, 62-66) In the Netherlands, patients with COPD are often treated in primary care physical therapy practice; for example, in 2017, a total of 35,227 patients with COPD received physical therapy treatment.(39). A Dutch clinical practice guideline was developed for patients with COPD, with meaningful outcome domains and associated measures that can be used as a basis for the further development of quality indicators.(67)

Objectives

The aim of this thesis is to develop, select and test core sets of outcome-based quality indicators in primary care physical therapy practice for patients with NSLBP and COPD, based on consensus between stakeholders. The selection of the core sets by stakeholders will be supported with routinely collected real-world data and include relevant outcome domains and associated measures to monitor and evaluate primary care physical therapy treatment for patients with NSLBP and COPD. The comparison of meaningful outcomes between physical therapists must be valid, reliable and accepted as having added value by the stakeholders. See box 1 for the research questions.

Box 1 Research questions

- What standard set of outcome domains and associated measures for patients with NSLBP can be developed in Dutch primary care physical therapy practice?
- What standard set of outcome domains and associated measures for patients with COPD can be developed in Dutch primary care physical therapy practice?
- Which potential outcome-based quality indicators for patients with NSLBP can be used for the selection of a core set based on the acceptance of stakeholders?
- Which potential outcome-based quality indicators for patients with COPD can be used for the selection a core set based on the acceptance of stakeholders?
- What are the experiences of physical therapists in the use of a standard set for patients with COPD in terms of the interaction between patients, quality-improvement initiatives and public reporting?

Thesis outline

Chapter 2 describes the results of a Delphi consensus study with stakeholders (patients, physical therapists, policy-makers and health insurers) for the development of a standard set of outcome domains and associated PROMs for patients with NSLBP in Dutch primary care physical therapy practices.

Chapter 3 reports on a Delphi consensus study with stakeholders (patients, physical therapists, policy-makers and health insurers) that describes the results of developing a standard set of outcome domains and associated measures for patients with COPD in Dutch primary care physical therapy practices.

Chapter 4 defines the potential outcome-based quality indicators for patients with NSLBP based on prospectively collected cohort data comprising PROMs. Focus groups were conducted with stakeholders (physical therapists and senior researchers) to select a core set of quality indicators that they perceived to add value as a quality-improvement tool.

Chapter 5 defines the potential outcome-based quality indicators for patients with COPD, supported by prospectively collected cohort data comprising PROMs and physical performance measures. Focus group were conducted with stakeholders (physical therapists and senior researchers) to select a core set of quality indicators perceived to add value as a quality-improvement tool.

Chapter 6 presents a study exploring the implementation of the set of measurement instruments for patients with COPD in physical therapy practice, including how the set can be used for goal setting, quality improvement and external transparency.

Lastly, in **chapter 7**, the general discussion, the results of all chapters are discussed and taken into a broader theoretical and practical perspective. This chapter also includes conclusions and recommendations for further research.

REFERENCES

1. Richardson W. Crossing the Quality Chasm; A new health system for the 21st century Institute of Medicine. 2001;(March), 1–8.
2. Scholte M, Neeleman-Van der Steen CW, Hendriks EJ, Nijhuis-Van der Sanden MW, Braspenning J. Evaluating quality indicators for physical therapy in primary care. International Journal for Quality in Health Care. 2014;26(3):261-70.
3. Scholte M, van Dulmen SA, Neeleman-Van der Steen CW, van der Wees PJ, Nijhuis-van der Sanden MW, Braspenning J. Data extraction from electronic health records (EHRs) for quality measurement of the physical therapy process: comparison between EHR data and survey data. BMC medical informatics and decision making. 2016;16(1):141.
4. Oostendorp RA, Rutten GM, Dommerholt J, Nijhuis-van der Sanden MW, Harting J. Guideline-based development and practice test of quality indicators for physiotherapy care in patients with neck pain. J Eval Clin Pract. 2013;19(6):1044-53.
5. Gijsbers HJ, Lauret GJ, van Hofwegen A, van Dockum TA, Teijink JA, Hendriks HJ. Development of quality indicators for physiotherapy for patients with PAOD in the Netherlands: a Delphi study. Physiotherapy. 2016;102(2):196-201.
6. Nijkrake MJ, Keus SH, Ewalds H, Overeem S, Braspenning JC, Oostendorp RA, et al. Quality indicators for physiotherapy in Parkinson's disease. Eur J Phys Rehabil Med. 2009;45(2):239-45.
7. Mant J. Process versus outcome indicators in the assessment of quality of health care. International Journal for Quality in Health Care. 2001;13(6):475-80.
8. Porter ME. What is value in health care? The New England journal of medicine. 2010;363(26):2477-81.
9. Van Der Wees PJ. Framework for the use of PROMs. In: Recommendations to OECD Ministers of Health from the high level reflection group on the future of health statistics. Paris: OECD. 2017;27.
10. Van Der Wees PJ, Nijhuis-Van Der Sanden MW, Ayanian JZ, Black N, Westert GP, Schneider EC. Integrating the use of patient-reported outcomes for both clinical practice and performance measurement: views of experts from 3 countries. The Milbank quarterly. 2014;92(4):754-75.
11. Hamilton AB, Yano EM. The importance of symbolic and engaged participation in evidence-based quality improvement in a complex integrated healthcare system: response to "The science of stakeholder engagement in research". Translational behavioral medicine. 2017;7(3):492-4.
12. Goodman MS, Sanders Thompson VL. The science of stakeholder engagement in research: classification, implementation, and evaluation. Translational behavioral medicine. 2017;7(3):486-91.
13. Campbell SM, Braspenning J, Hutchinson A, Marshall MN. Research methods used in developing and applying quality indicators in primary care. BMJ. 2003;326(7393):816-9.
14. Cella D, Hahn EA, Jensen SE, Butt Z, Nowinski CJ, Rothrock N, editors. Methodological issues in the selection, administration and use of patient-reported outcomes in performance measurement in health care settings. NQF (NQF), Washingtdon, DC: National Quality Forum (NQF); 2012.
15. International Consortium For Health Outcome Measurement (ICHOM) [Available from: https://www.ichom.org/.

16. Berwick DM, Black N, Cullen D, Deerberg-Wittram J, Diverty B, Epstein A, et al. Recommendations to OECD Ministers of Health from the high level reflection group on the future of health statistics. In: OECD P, editor. 2017.

17. Olde Rikkert MGM, van der Wees PJ, Schoon Y, Westert GP. Using Patient Reported Outcomes Measures to Promote Integrated Care. Int J Integr Care. 2018;18(2):8-.

18. Porter ME, E.O. T. Redefining health care: creating value-based competition on results.: Boston: Harvard Business School Press; 2006.

19. Development of outcome focused care. In: Ministry of Health WaSVitN, editor.

20. Santeon; hospital network in the Netherlands [Available from: https://santeon.nl/over-santeon/missie/.

21. Meerhoff GA, van Dulmen SA, Maas MJM, Heijblom K, Nijhuis-van der Sanden MWG, Van der Wees PJ. Development and Evaluation of an Implementation Strategy for Collecting Data in a National Registry and the Use of Patient-Reported Outcome Measures in Physical Therapist Practices: Quality Improvement Study. Physical Therapy. 2017;97(8):837-51.

22. The national data registry (LDK) of the Association for Quality in Physical Therapy (SKF) [Available from: https://www.keurmerkfysiotherapie.nl/praktijk/data/.

23. Patient Reported Outcomes (PROs) in Performance Measurement Washington, DC: National Quality Form (NQF); 2012 [Available from: http://www.qualityforum.org/Publications/2012/12/Patient-Reported_Outcomes_in_Performance_Measurement.aspx.

24. Westby MD, Klemm A, Li LC, Jones CA. Emerging role of quality indicators in physical therapist practice and health service Delivery. Physical therapy. 2016;96(1):90-100.

25. van der Wees PJ, Verkerk EW, Verbiest MEA, Zuidgeest M, Bakker C, Braspenning J, et al. Development of a framework with tools to support the selection and implementation of patient-reported outcome measures. Journal of Patient-Reported Outcomes. 2019;3(1):75.

26. Weldring T, Smith SM. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). Health Serv Insights. 2013;6:61-8.

27. Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, et al. The Quebec Back Pain Disability Scale. Measurement properties. Spine. 1995;20(3):341-52.

28. Jette DU, Halbert J, Iverson C, Miceli E, Shah P. Use of standardized outcome measures in physical therapist practice: perceptions and applications. Phys Ther. 2009;89(2):125-35.

29. Westby MD, Marshall DA, Jones CA. Development of quality indicators for hip and knee arthroplasty rehabilitation. Osteoarthritis and cartilage. 2018;26(3):370-82.

30. Fitch K, Steven J. Bernstein, Maria Dolores Aguilar, Bernard Burnand, Juan Ramon LaCalle, Pablo Lazaro, Mirjam van het Loo, Joseph McDonnell, Janneke Vader and James P. Kahan. The RAND/UCLA Appropriateness Method User's Manual. RAND Corporation. 2001.

31. Kötter T, Blozik E, Scherer M. Methods for the guideline-based development of quality indicators--a systematic review. Implementation Science. 2012;7(1):21.

32. Langendam MW, Piggott T, Nothacker M, Agarwal A, Armstrong D, Baldeh T, et al. Approaches of integrating the development of guidelines and quality indicators: a systematic review. BMC health services research. 2020;20(1):875.

33. Oostendorp RAB, Hans Elvers JW, van Trijffel E. The quality of physiotherapy care: the development and application of quality indicators using scientific evidence and routinely collected data embedded in the process of clinical reasoning. Bulletin of Faculty of Physical Therapy. 2019;24(2):113-20.

34. van Weel C, Schers H, Timmermans A. Health care in the Netherlands. J Am Board Fam Med. 2012;25 Suppl 1:S12-7.

35. Wammes J, Stadhouders N, Westert GP. International Health Care System Profiles: Netherlands [Available from: https://www.commonwealthfund.org/international-health-policy-center/countries/netherlands.

36. Equalis E. Public explanatory document national treatment index 2018.

37. Skipr. Fysiotherapeut is bureaucratie beu [Available from: https://www.skipr.nl/blog/fysiotherapeut-is-bureaucratie-beu/.

38. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, et al. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society. 2006;15 Suppl 2:S192-300.

39. Vektis [Available from: https://www.zorgprismapubliek.nl/producten/zorg-dichtbij/fysiotherapie/row-5/voor-welke-diagnoses-binnen-de-basisverzekering-komen-mensen-het-meest-bij-de-fysiotherapeut/.

40. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global Strategy for the Diagnosis, Management and Prevention of Chronic Obstructive Lung Disease 2017 Report: GOLD Executive Summary. Respirology (Carlton, Vic). 2017;22(3):575-601.

41. Lennon S, Hastings M. Key Physiotherapy Indicators for Quality of Stroke Care. Physiotherapy. 1996;82(12):655-64.

42. Peter WF, Hurkmans EJ, Wees PJvd, Hendriks EJM, Bodegom-Vos Lv, Vlieland TPMV. Healthcare Quality Indicators for Physiotherapy Management in Hip and Knee Osteoarthritis and Rheumatoid Arthritis: A Delphi Study. Musculoskeletal Care. 2016;14(4):219-32.

43. Petersson IF, Strombeck B, Andersen L, Cimmino M, Greiff R, Loza E, et al. Development of healthcare quality indicators for rheumatoid arthritis in Europe: the eumusc.net project. Ann Rheum Dis. 2014;73(5):906-8.

44. Scholte M, Neeleman-van der Steen CW, Hendriks EJ, Nijhuis-van der Sanden MW, Braspenning J. Evaluating quality indicators for physical therapy in primary care. International journal for quality in health care : journal of the International Society for Quality in Health Care. 2014;26(3):261-70.

45. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, et al. What low back pain is and why we need to pay attention. Lancet (London, England). 2018;391(10137):2356-67.

46. Burton AK, Balagué F, Cardon G, Eriksen HR, Henrotin Y, Lahad A, et al. Chapter 2 European guidelines for prevention in low back pain: November 2004. European Spine Journal. 2006;15(Suppl 2):s136-s68.

47. Van den Dool J ST. Nivel Zorgregistraties – Zorg door de fysiotherapeut; jaarcijfers 2018 en trendcijfers 2014 – 2018 2019 [Available from: https://www.nivel.nl/sites/default/files/bestanden/1003595.pdf.

48. Foster NE, Anema JR, Cherkin D, Chou R, Cohen SP, Gross DP, et al. Prevention and treatment of low back pain: evidence, challenges, and promising directions. Lancet (London, England). 2018;391(10137):2368-83.

49. Rutten GM, Degen S, Hendriks EJ, Braspenning JC, Harting J, Oostendorp RA. Adherence to Clinical Practice Guidelines for Low Back Pain in Physical Therapy: Do Patients Benefit? Physical Therapy. 2010;90(8):1111-22.

50. Scheele J, Vijfvinkel F, Rigter M, Swinkels IC, Bierman-Zeinstra SM, Koes BW, et al. Direct access to physical therapy for patients with low back pain in the Netherlands: prevalence and predictors. Phys Ther. 2014;94(3):363-70.

51. Swinkels IC, Kooijman MK, Spreeuwenberg PM, Bossen D, Leemrijse CJ, van Dijk CE, et al. An overview of 5 years of patient self-referral for physical therapy in the Netherlands. Phys Ther. 2014;94(12):1785-95.

52. Bekkering GE, van Tulder MW, Hendriks EJ, Koopmanschap MA, Knol DL, Bouter LM, et al. Implementation of Clinical Guidelines on Physical Therapy for Patients With Low Back Pain: Randomized Trial Comparing Patient Outcomes After a Standard and Active Implementation Strategy. Physical Therapy. 2005;85(6):544-55.

53. Bekkering GE, Hendriks HJ, van Tulder MW, Knol DL, Hoeijenbos M, Oostendorp RA, et al. Effect on the process of care of an active strategy to implement clinical guidelines on physiotherapy for low back pain: a cluster randomised controlled trial. Quality & safety in health care. 2005;14(2):107-12.

54. Staal J, Hendriks E, Heijmans M, Kiers H, Lutgers-Boomsma G, Rutten G. KNGF-richtlijn Lage rugpijn Verantwoording en toelichting. KNGF. 2013;21.

55. WHO. World Health Statistics 2008. Switzerland: World Health Organization. 2008.

56. Gershon AS, Warner L, Cascagnette P, Victor JC, To T. Lifetime risk of developing chronic obstructive pulmonary disease: a longitudinal population study. Lancet (London, England). 2011;378(9795):991-6.

57. WHO. Key facts of Chronic obstructive pulmonary disease (COPD) 2017 [Available from: https://www.who.int/en/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd).

58. Rochester CL, Vogiatzis I, Holland AE, Lareau SC, Marciniuk DD, Puhan MA, et al. An Official American Thoracic Society/European Respiratory Society Policy Statement: Enhancing Implementation, Use, and Delivery of Pulmonary Rehabilitation. American journal of respiratory and critical care medicine. 2015;192(11):1373-86.

59. McCarthy B, Casey D, Devane D, Murphy K, Murphy E, Lacasse Y. Pulmonary rehabilitation for chronic obstructive pulmonary disease. The Cochrane database of systematic reviews. 2015(2):Cd003793.

60. Spruit MA, Van't Hul A, Vreeken HL, Beekman E, Post MHT, Meerhoff GA, et al. Profiling of Patients with COPD for Adequate Referral to Exercise-Based Care: The Dutch Model. Sports Medicine. 2020;50(8):1421-9.

61. Koolen EH, van Hees HW, van Lummel RC, Dekhuijzen R, Djamin RS, Spruit MA, et al. "Can do" versus "do do": A Novel Concept to Better Understand Physical Functioning in Patients with Chronic Obstructive Pulmonary Disease. J Clin Med. 2019;8(3):340.

62. Spruit MA, Singh SJ, Garvey C, ZuWallack R, Nici L, Rochester C, et al. An official American Thoracic Society/European Respiratory Society statement: key concepts and advances in pulmonary rehabilitation. American journal of respiratory and critical care medicine. 2013;188(8):e13-64.

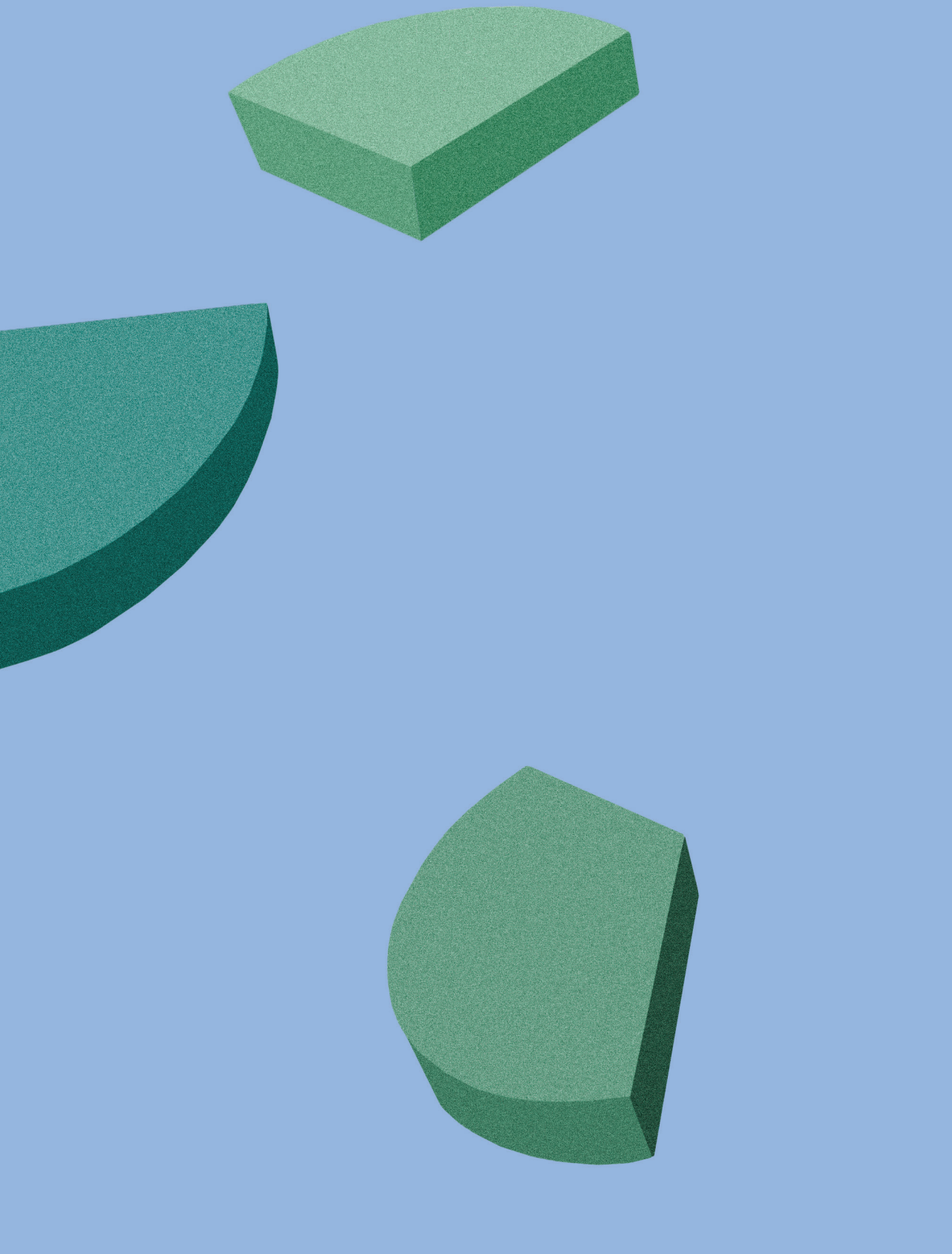
63. Puhan MA, Gimeno-Santos E, Scharplatz M, Troosters T, Walters EH, Steurer J. Pulmonary rehabilitation following exacerbations of chronic obstructive pulmonary disease. The Cochrane database of systematic reviews. 2011(10):Cd005305.

64. Dimitrova A, Izov N, Maznev I, Vasileva D, Nikolova M. Physiotherapy in Patients with Chronic Obstructive Pulmonary Disease. Open Access Maced J Med Sci. 2017;5(6):720-3.

65. Hul Avt, Vreeken H, Gosselink R, Langer D, Meerhoff G, Muris J, et al. Fysiotherapie bij COPD helpt, maar blijft vaak onbenut. Huisarts en wetenschap. 2019;62(10):34-7.

66. Beekman E, Mesters I, Hendriks E, Muris J, Wesseling G, Evers S, et al. Exacerbations in patients with chronic obstructive pulmonary disease receiving physical therapy: A cohort-nested randomised controlled trial. BMC pulmonary medicine. 2014;14:71.

67. Gosselink R, Langer D, Burtin C, Probst V, Hendriks HJM, Schans CPvd, et al. KNGF-Guideline for physical therapy in patients with chronic obstructive pulmonary diseasehttps://www.fysionet-evidencebased.nl/images/pdfs/richtlijnen/copd_2008/copd_verantwoording_en_toelichting.pdf. 2008.



CHAPTER 2

Development of a standard set of outcome measures for nonspecific low back pain in Dutch primary care physiotherapy practices

A Delphi study

Arie C. Verburg
Simone A. van Dulmen
Henri Kiers
Maria W.G. Nijhuis-van der Sanden
Philip J. van der Wees

Published in European Spine Journal volume 28, pages1550–1564 (2019)
Digital object identifier: 10.1007/s00586-019-05962-x

ABSTRACT

Purpose To develop a standard set of outcome measures that are accepted for relevance and feasibility by stakeholders and useful for a) interaction between patient and the professional, e.g. shared decision making in goal-setting, monitoring and feedback based on outcomes b) internal quality improvement, and c) external transparency in patients with nonspecific low back pain (NSLBP) in primary care physical therapy.

Methods We used a consensus-driven modified RAND-UCLA Delphi method in seven steps with panellists (patients, representatives of patient- and physiotherapy associations, researchers, policy makers, health insurers) (1) literature search, (2) first online survey, (3) patient interviews, (4) an experts meeting, (5) a consensus meeting, (6) second online survey, and (7) final approval of an advisory board. Steps 1-4 resulted in potential outcome measures. In the consensus meeting after discussion panellists voted for inclusion per measure. In the second online survey the final standard set was rated on relevance and feasibility on a 9-point Likert scale, when the median score was ≥7 the standard set was accepted and finally approved.

Results Thirteen draft outcome measures were rated and discussed, and finally six outcome measures were accepted. The standard set includes the Quebec Back Pain Disability Scale (QBPDS), Oswestry Disability Index (ODI), Patient Specific Functional Scale (PSFS), Numeric Pain Rating Scale (NPRS), Global Perceived Effect (GPE-DV), and the Start Back Screening Tool (SBT)**.**

Conclusion This study presents a standard set of outcome measures for patients with NSLBP in primary care physiotherapy accepted for relevance and feasibility by stakeholders. The standard set is currently used in daily practice and tested on validity and reliability in a pilot study.

INTRODUCTION

Quality of care has been defined as: *"doing the right thing, at the right time, in the right way, for the right person, and having the best possible results*[1]. The challenge is how to measure quality of care in daily practice. Results at the level of the patient's health status can be measured with patient reported outcome measures (PROMs), patient reported experience measures (PREMs) and/or physical performance measures. PROMs are questionnaires or single-item scales measuring aspects of a patient's health status directly reported by the patient, e.g. perceived pain. PREMs are questionnaires measuring the experience of patients with healthcare, e.g. communication with the healthcare professional. Physical performance measures are clinical tests to measure physical function, e.g. 6-minute walking test. Outcome measures should be well developed and unidimensional in order to generate information regarding the construct of interest.[2] Using such combined outcomes in daily practice is proposed to facilitate the interaction between patient and healthcare professional, including the process of shared decision making, goal setting, and evaluation of treatment effects [3,4].

Outcomes measurement can also be useful to provide transparency about the process and the intervention effect at the group level in order to facilitate quality improvement trajectories, to provide information for patients, and for pay for performance purposes [5-7]. For successful implementation of outcome measures, patients and healthcare professionals need to accept a common set of outcomes to be measured as having added value in daily practice. In physiotherapy practice, multiple outcome measures are being used in clinical decision making. Routine data collection in daily practice opens the opportunity for establishing large data sets with patient outcomes. Standardization of these measurements is necessary to enable comparison of intervention effects [3,8].

Currently, limited data is available about the quality of daily care of patients with nonspecific low back pain (NSLBP) treated in physiotherapy practices in the Netherlands. Therefore, this study focuses on the development of a standard set of outcome measures for NSLBP, the most common health condition of patients visiting physiotherapy in primary care practice. The final set of measures should be accepted as having added value in clinical practice and will have to be useful to compare the outcomes at the level of the individual patient, and for measuring and improving quality of physiotherapists and their practices.

Previous international studies showed several initiatives for developing outcome sets for low back pain [9-14]. Most of these standard outcome sets were developed for clinical trial purposes, and have

not been tested with regards to relevance and feasibility in the evaluation of quality of care in daily practice. Few studies on NSLBP in physiotherapy showed a good relationship of higher guideline adherence with better outcomes and less utilization of care [15,3]. This stresses the value of gaining insight into outcomes on a larger scale. Successful implementation of outcome measures in daily practice can be improved by stakeholder engagement in quality improvement initiatives [16,17]. It is therefore important to include all relevant stakeholders in Dutch physiotherapy concerning NSLBP in the development of the current standard set of outcomes.

In this study, NSLBP was defined as pain and discomfort, localized below the costal margin and above the inferior gluteal folds, with or without leg pain, and not caused by specific pathology [18]. NSLBP is an example of a patient group with high variation in level of recovery and time to recover varying from one day to multiple years. There is an increasing popularity for stratification of patients with NSLBP in subgroups, taking into account differences in characteristics based on prognostic profiles [19-22]. Outcome sets combined with stratified care will be more precise and useful for the development of quality indicators and better accepted for quality improvement [23].

The aim of this study was to develop a clinical standard set of outcome measures in patients with NSLBP - taking into account classification in clinically relevant subgroups - that is accepted for relevance and feasibility by stakeholders, and deemed useful for a) interaction between patient and healthcare professional, e.g.shared decision making in goalsetting and monitoring and feedback based on outcomes b) internal quality improvement, and c) external transparency of primary care physical therapist practices.

METHODS

Design

This study used a mixed method design in Dutch physiotherapy practices. The study was conducted between October 2016 and July 2017. An advisory board was formed with representatives of the Dutch Patient Association for Back Pain (NVVR), the Royal Dutch Society for Physiotherapy (KNGF), the Association for Quality in Physiotherapy (SKF) and two representatives of health insurance companies in the Netherlands (CZ & DFZ) to monitor and evaluate the process and to facilitate the implementation by providing and receiving information from stakeholder groups.

We used a consensus-driven RAND/UCLA modified Delphi method to select relevant outcomes [24] in seven separate steps (see table 1). Informed consent was obtained from all individual participants included in the study and all procedures were conducted according to the Declaration of Helsinki

Table 1 Steps during the consensus-driven modified Delphi method

Step	Participants	Goal	Result
1. Explorative review of the literature	Research team (AV, SD, RN, and PW)	Identification of valid structure-, process-, and outcome (PROMs, PREMs, physical performance measures) measures for NSLBP that are described in scientific literature. Screening for profiles in subgroups related to the course of recovery based on prognostic factors	Selection and analysis of all eligible structure-, process-, and outcome measures on their validity and reliability. Selection of eligible prognostic profiles for NSLBP.
2. First online survey round	32 out of 43 Dutch physiotherapists with ample treatment or scientific experience on NSLBP	Anonymous rating of structure, process and outcome measures with a 9-point Likert scale on relevance and feasibility.	The median score of every measure was calculated.
3. Expert committee	4 senior physiotherapists of the first online survey round	Results of the online survey round 1 was interpreted and discussed with experts	First interpretation of the results of the first online survey presented and discussed
4. Interview with patients	6 patients with NSLBP and treated by physiotherapists last year	Patients views of measurements in clinical practice	The patients' perspectives on the use of measurement instruments which were presented during the introduction of the consensus meeting (step 5)
5. Consensus meeting	16 out of 43 participants of the first online survey and representatives of patients, KNGF, SKF, the advisory board and policy makers	Nominal group technique was used to discuss the draft set. Finally, the participants voted on feasibility and relevance.	Measures were included if 80% or higher voted yes for inclusion. Between 60 – 80 % yes were deferred and rerated in the second online survey and between 0 and 60 % yes- were excluded.
6. Second online survey round	29 out of 43 participants of the first online survey (step 2)	The participants rated the second draft set and if needed alternatives on a 9-point Likert scale	Measures were included when all participants rated a median of 7 or higher on the 9-point Likert scale.
7. Final approval of the advisory board	Advisory board (MJS, LV, HW, AW, AT and ML)	The advisory board was asked to accept the final set	Final acceptance of the set

Step 1. Explorative review of the literature

We searched in the Guideline International Network (G-I-N)- and PEDro database for outcomes measures based on PROMs, PREMs and physical performance measures with adequate psychometric properties, including reliability, validity and responsiveness. [25,26] All multi- and mono-disciplinary Dutch and international

clinical practice guidelines for physiotherapists, general practitioners, and medical specialists were included. We also searched websites of organizations developing clinical practice guidelines (see appendix A). Based on the identified guidelines, reference tracking was performed. We preferred outcome measures that were already used in daily practice in the Netherlands. We also searched for structure, process and outcome measures in existing indicator sets, see appendix B for the search string. We used a pragmatic explorative approach and did not aim at conducting a systematic review of the literature. In the next step we analysed all eligible measures on their validity and reliability. The following information was gathered: type of measure (process, structure, outcome), type of questionnaire/instrument, targeted patient group, content of the questionnaire/instrument, time to complete the questionnaire/instrument, the minimal clinically important difference (MCID), domain (e.g. pain), related measurements, whether the questionnaire/instrument was already translated in Dutch, and supporting literature.

We used the prognostic profiles of the KNGF guideline as primary classification for subgroups related to the course recovery based on prognostic factors [27]. Then, we used the same literature search as for the outcome measures to compare the prognostic profiles of the KNGF with multi- and mono-disciplinary Dutch and international clinical practice guidelines. If necessary, we combined useful elements of different guidelines with the profiles of the KNGF guideline. In the identified guidelines, reference tracking was performed. Additionally, the PubMed database was screened between January 2012 and December 2016 for systematic reviews about individual prognostic factors in NSLBP, see appendix B for the search string. The individual prognostic factors were used as addition on the search to prognostic profiles. After the screening we selected all useful factors for prognostic profiles to classify subgroups for NSLBP.

Step 2. First online survey round

We recruited 43 Dutch physiotherapists via contact networks of the KNGF and SKF for participation in the Delphi rounds, this was a purposive sample. All participants needed to have ample treatment experience in patients with NSLBP, or experience in scientific research on NSLBP, or both. The goal of this step was to rate all in step 1 selected measures with a 9-point Likert scale on relevance and feasibility. Afterwards the participants rated the appropriateness and feasibility of prognostic profiles for NSLBP. We conducted the online survey in LimeSurvey version 2.06.

Step 3. Expert committee

We invited four participants with complementary expertise of the online survey to join an expert committee to discuss the results of the online survey and rated the outcome set and prognostic profiles on its content validity and reliability.

Step 4. Patient interviews

We invited six patients who were treated by physiotherapists for NSLBP in the past year. The patients were recruited via a convenience sample of six physiotherapy practices. Each physiotherapy practice included one adult patient that was treated for NSLBP in the last year. When the patient agreed on the informed consent, the physiotherapist gave the contact information to the researcher. Short semi-structured telephone interviews of approximately 30 minutes were held by two researchers (AV, JL) and a topic list was used. The aim of the interviews was to gain insight into the patient perspective on relevance and feasibility on the use of questionnaires (PROMs and PREMs), physical performance measures, process- and structure measures, and to what extent measurements can be used for improving the quality of care. The interviews were audio-recorded, transcribed verbatim and analysed using thematic analysis. AV and JL independently analysed the interviews and assigned codes within and between the interviews. Afterwards the assigned codes were compared and grouped together in greater categories and themes. The most important themes were presented during the consensus meeting.

Step 5. Consensus meeting

All participants of the online survey were invited in a three-hour consensus meeting, together with policy makers and members of the advisory board. We used the Nominal Group Technique (NGT) to structure the meeting [28,29]. The NGT is specifically designed and widely used for consensus statements between experts in a certain topic [28,29]. The different steps in the NGT helps to give all participants a voice in the consensus process [30]. During these steps the participants rated, discussed, and then re-rated the eligible structure-, process- and outcome measures. We discussed all measures scored in the first online survey and presented the results of step 2-4, followed by a second onsite rating of the relevance and feasibility of the measures. The measures were included in the standard set if the total votes scored 80% or higher on yes/no rating. All measures that were scored between 60% and 80% in the onsite rating were deferred for discussion in the second online survey. All measures that received between 0% and 60% of the votes were excluded.

Step 6. Second online survey round

All participants were invited for the second online round. All outcome measures that received between 60% and 80% of the votes were re-rated on relevance, and if needed, alternatives that were discussed in step 5 were rated. Based on these results, we developed a final outcome set. We concluded the set and the prognostic profiles as being accepted when all panellists rated a median of 7 or higher on a 9-point Likert scale.

Step 7. Final approval of the advisory board
Finally, the advisory board was asked to accept the final outcome set. The goal was to inform every stakeholder about the conclusions and implications of the study and to increase the acceptability. All representatives were asked to take the responsibility for communication of the results of the study in their own organisation.

RESULTS

Participants
In step 2, 32 of the 43 panellists (response rate: 70%) completed the survey. The mean age of the panellists was 42 years and 75.8% were men. In step 3, four expert physiotherapists (FM, DH, BM, BH) accepted to participate. During step 4, six semi-structured interviews with six patients with NSLBP were held. The age ranged from 42 – 73 years with an average age of 56; four of the participants were men. In step 5, the consensus meeting, 16 of the 43 physiotherapists and researchers participated (response rate: 37%), as well as seven policy makers and members of the advisory board. The patient representatives were not able to join at the meeting. For the second online survey in step 6, 29/43 (response rate: 68%) respondents participated. Finally in step 7, the five members of the advisory board participated.

Step 1. Explorative review of the literature
We identified 27 measures of which 13 were eligible for further investigation: six PROMs, two PREMs, two additional outcome measures two process measures and one screening tool (see table 2).

The reasons for exclusion of the 14 measures were: not familiar in the Netherlands, not developed and useful for NSLBP, not primarily advised by guidelines, and not for physiotherapy primary care purposes. For a clear description of all included measures and an overview of all excluded measures, see appendix C.

Development of prognostic profiles
We identified 19 guidelines, of which ten described useful information about prognostic profiles for NSLBP or individual prognostic factors.[27,21,31-39] The remaining nine guidelines were focused on specific pathology. To develop prognostic profiles we focused primarily on the Dutch KNGF guideline and compared it with other guidelines.[27] The majority of the guidelines specified two or three prognostic patient profiles based on the expected time of recovery. Some guidelines did not provide prognostic profiles in a table, but described them narratively.[35-37] Furthermore, all guidelines described individual prognostic factors that are associated with the course of recovery. See appendix D for a summary of all useful components.

Based on the outcomes of the literature review, we distinguished three prognostic profiles for NSLBP (A, B, C), all profiles containing four characteristics. These characteristics are generally based on prognostic (psychosocial) factors of the KNGF guideline. Some guidelines described the expected time of recovery in weeks; this was added to our prognostic profiles. We identified prognostic factors based on: back pain-related factors, individual factors, work-related factors and psychosocial factors. The profiles are described in Appendix E.

Step 2. First online survey round

Of the 13 outcome measures that were rated, five outcome measures scored a median of ≥7 on *relevance* and on *feasibility* nine outcome- and process measures scored a median of ≥7. The prognostic profiles scored a median of ≥*7* on *relevance* and *feasibility***;** See table 2 for a more specified overview of the survey.

Table 2 Result of step 2: first online survey and step 5: consensus meeting

Instrument/process measures	**Type of measure**	**Relevance (step 2)**	**Feasibility (step 2)**	**Inclusion for the final standard set (step 5)[b]**	
		Median	Median	Yes (%)	No (%)
Quebec Back Pain Disability Scale (QBPDS)	Outcome	7	7	80[c]	20
Oswestry Disability Index (ODI)	Outcome	5	7	64[b]	36
Patient Specific Functional Scale (PSFS)	Outcome	7	8	76[b]	24
Numeric Pain Rating Scale (NPRS)	Outcome	7	9	100[c]	0
Global Perceived Effect Dutch Version (GPE-DV)	Outcome	7	8	88[c]	12
Improvement of activities	Outcome	6	7	0	100
Patient Reported Experience Measures (PREM)	Outcome	7	5	19	81
Subgroups for Targeted Treatment (STarT) Back Screening Tool (SBT)	Screening Tool	4	7	93[c]	7
Tampa Scale of Kinesiophobia (TSK)	Outcome	4	6	0	100
Number of treatment sessions	Outcome	5	8	Not rated[a]	
History taking described in the EHR	Process	6	7	0	100
Treatment plan described in the EHR	Process	6	7	0	100
Prognostic profiles from guidelines	Prognostic profiles	7	7	Not rated[a]	

[a]The number of treatment sessions and prognostic profles was not rated because the panellists suggested an alternative instrument
[b]Re-rated in the second online survey
[c]Final inclusion in the standard set after rating
EHR Electronic Health Record

Step 3. Expert committee

During the expert meeting all 13 selected outcome and process measures were discussed for validity and reliability. The experts accepted the prognostic profiles as having added value in daily practice. They made some suggestions, as to their opinion prognostic profiles are still not accurate enough to predict individuals who develop chronic pain or not. For example, prognostic factors of a patient in profile A, can also be seen in profile B. Acute, sub-acute and chronic low back pain should not be used in the prognostic profiles. The experts suggested selecting outcome measures per profile. The experts stated that it will be necessary to perform a solid pilot study to test the selected outcome and process measures on feasibility before the outcome set can be considered valid and reliable for quality improvement purposes.

Step 4. Patient interviews

The following themes were identified: 1) patient satisfaction, 2) administration, 3) number of treatment sessions, 4) transparency and 5) PROMs. Almost every patient agreed that satisfaction about the given treatment and treatment effect is relevant for quality evaluation purposes. Clinical record keeping is important to monitor the effect of treatment, to support a colleague during takeovers, and is valuable for evaluating quality of care, but should also be short and brief. The number of treatment sessions can be useful to evaluate quality of care, depending on the patient group. Some patients stated that transparency about outcomes of care could help them to choose healthcare professionals and other patients said that they preferred the advice of a doctor, therapist or family member. There were different opinions whether PROMs were relevant for quality evaluation purposes. Most patients stated that the readability of PROMs is good. Some patients said that pain and functional problems were useful elements to score in PROMs but also psychosocial factors as this may have influence on the effect of treatment. See appendix F for the themes, categories and codes.

Step 5. Consensus meeting

After presentation of the results of previous steps, the panellists could vote per measure with yes/no whether it should be added to the final outcome set; all results are presented in table 2. The Quebec Back Pain Disability Scale (QBPDS), Numeric Pain Rating Scale (NPRS), Global Perceived Effect- Dutch Version (GPE-DV) and Start Back Screening Tool (SBT) were chosen to include to the final outcome set directly. The Oswestry Disability Index (ODI), Patient Specific Functional Scale (PSFS), number of treatment sessions and prognostic profiles were scored between 60-80% and the following suggestions were done during the consensus meeting for adapting the measures: The panellists suggested that the ODI is less common in the Netherlands compared to the QPBDS. Nevertheless, the ODI is widely used internationally and consists of good psychometric properties. The panellists suggested adding the ODI to the outcome set and compare the ODI with the QBPDS in a pilot study. The pilot study opens the opportunity to reflect on preferences of the

field and to analyse the feasibility, acceptability and responsiveness between both instruments. In the meantime, physiotherapists in the field can choose between the ODI and QBPDS. The panellists stated that the number of treatment sessions should be changed to measuring the total costs of the episode. The panellists suggested that the start back screening tool (SBT) could replace the prognostic profiles.

Step 6. Second online survey round

The following four questions related to the remaining measures were rated on a 9-point Likert scale: 'Do you agree that we use the SBT to classify patients in subgroups?', 'Do you agree that we add the PSFS in the outcome set?', 'Do you agree that the ODI will be tested in a pilot study in comparison with the QBPDS?', and 'Do you agree that treatment costs should be evaluated but will not be added in the outcome set?' In the online survey we presented the final outcome set and prognostic profiles as in table 3. All questions scored a median of 7 or higher and therefore the outcome set was accepted.

Step 7. Final approval of the advisory board

The final set with outcome measures was accepted by the advisory board. They accepted the outcome set by signing an official approval document.

Table 3 Final set of measures accepted by all stakeholders

Profiles based on SBT	**Low risk profile[b]**		**Medium/high risk profile[b]**		
Measurement	**Intake[a]**	**End of treatment[a]**	**Intake[a]**	**Every six weeks[a]**	**End of treatment[a]**
Quebec Back Pain Disability Scale (QBPDS)			X	X	X
Oswestry Disability Index (ODI)(pilot)			X	X	X
Patient Specific Functional Scale (PSFS)	X	X	X	X	X
Numeric Pain Rating Scale (NPRS)	X	X	X	X	X
Global Perceived Effect Dutch Version (GPE-DV)		X			X

[a]Intake, every six weeks and end of treatment refer to the time that physiotherapists need to score the outcomes

[b]The Subgroups for Targeted Treatment (STarT) Back Screening Tool (SBT) is used for allocating patients in low-, medium- or high-risk profile

DISCUSSION

This study presents a standard set of six clinical outcome measures in patients with NSLBP in primary care physiotherapy, which includes the Quebec Back Pain Disability Scale (QBPDS), Oswestry Disability Index (ODI), Patient Specific Functional Scale (PSFS), Numeric Pain Rating Scale (NPRS), Global Perceived Effect- Dutch Version (GPE-DV), and the Start Back Screening Tool (SBT). The outcome measures are aimed to be used for the interaction between patient and physiotherapist, internal quality improvement, and for external transparency.

In our study, the Start Back Screening Tool was selected to allocate patients in subgroups, while research showed that cautiousness is required with respect to interpretation of prognostic tools.[20] In line with the conclusions of this study, Karran et al. (2017) concluded that prognostic screening instruments in primary care scored poorly at assigning higher risk scores to individuals who develop chronic pain, than those who will not.[20] However, other research showed that identifying subgroups of patients with NSLBP is still promising for future healthcare.[40] Multiple researchers support this vision.[22,20,21] During a pilot study, we should test whether the Start Back Screening Tool is reliable and valid for classifying patients in subgroups.

The fundamental difference with existing outcome sets for low back pain is that this outcome set is accepted by stakeholders as having added value in daily care. [9,14,10,12,13,11] Therefore, this new standard set provides a more promising basis for the implementation of quality indicators in clinical practice. Stakeholder engagement is essential for successful implementation of quality improvement initiatives. In comparison with traditional Delphi methods we performed additional activities to reach consensus based on the RAND/UCLA appropriateness method. Along with the anonymous online surveys and consensus meeting, we conducted an expert meeting, interviewed patients, and consulted an advisory board. With these steps, stakeholders were encouraged to use this outcome set in daily practice. In our study the patients were included to reflect on the selection of measures for the standard set and not as a separate qualitative study. The interviews were limited to six patients and we did not reached data saturation, which could lead to bias. However, the interviews gained sufficient insights of the patients' views of measurements in clinical practice.

During the consensus study we found that the stakeholders and physiotherapists and other showed a positive attitude about developing the standard set and its described goals. This may not be representative for the total population of physiotherapists in the Netherlands. The panellists of this study may have been early adopters and open for quality improvement or external transparency. During implementation of this standard set we will need to anticipate that physiotherapists need more information

about the benefits on standardisation of outcome measurements to provide insight in and to compare intervention effects. Implementation strategies may need to be aimed at knowledge, skills and attitudes of physiotherapists.

Due to pragmatic reasons we were not able to let panellists rate the outcome measures on a 9-point Likert scale during the consensus meeting, as preferred by the RAND-UCLA method.[24] The panellists voted with yes/no. Potentially this may have influenced the voting and panellists could feel peer pressured with the used method. We verified whether the panellists felt comfortable about the procedure, and they agreed and felt safe to give their opinion. We do not expect that the results would have been different when using a Likert scale.

Before implementation of a standard set in daily practice, it is important to develop an infrastructure for collection of the data.[3,8] For example, the standard set must be implemented in the Electronic Health Record (EHR) that physiotherapists use for their clinical record keeping.[3] This EHR must be connected to a secure central database before the outcomes can be analysed. Also the infrastructure must allow the possibility to give practices and physiotherapist feedback on outcomes and useful for quality improvement.

In this study we present a consensus-based standard set of outcome measures that is accepted for relevance and feasibility by stakeholders. Therefore, this standard outcome set provides a promising basis for further development of quality indicators in physiotherapy practice.[41] The standard set is currently used in daily practice and tested for validity and reliability in a pilot before it can be used for the development of quality indicators.[7] All stakeholders should stay engaged during further implementation of the standard set.

We acknowledge research members: Joan Luites (JL) and Juliette Cruijsberg (JC); expert committee: Francois Maissan (FM), David den Hartog (DH), Bob van den Meiracker (BM), and Marcel Heijmans (MH); advisory board: Annemarie Trompert (AT), Marije de Leur (ML), Leen Voogt (LV), Hans Wattel (HW), and Alida Wolters (AW).

REFERENCES

1. Richardson W (2001) Crossing the Quality Chasm; A new health system for the 21st century Institute of Medicine (March), 1–8
2. Streiner DL, Norman GR, Cairney J (2015) Health measurement scales: a practical guide to their development and use. Oxford University Press, USA,
3. Meerhoff GA, van Dulmen SA, Maas MJM, Heijblom K, Nijhuis-van der Sanden MWG, Van der Wees PJ (2017) Development and Evaluation of an Implementation Strategy for Collecting Data in a National Registry and the Use of Patient-Reported Outcome Measures in Physical Therapist Practices: Quality Improvement Study. Physical Therapy 97 (8):837-851.
4. Reuben DB, Tinetti ME (2012) Goal-oriented patient care--an alternative health outcomes paradigm. The New England journal of medicine 366 (9):777-779.
5. van der Wees PJ, Zaslavsky AM, Ayanian JZ (2013) Improvements in Health Status after Massachusetts Health Care Reform. The Milbank quarterly 91 (4):663-689
6. Van Der Wees PJ, Nijhuis-Van Der Sanden MW, Ayanian JZ, Black N, Westert GP, Schneider EC (2014) Integrating the use of patient-reported outcomes for both clinical practice and performance measurement: views of experts from 3 countries. The Milbank quarterly 92 (4):754-775.
7. Verkerk EV, M. van Dulmen, S. van der Wees, P. Terwee, C. Beurskens, S. de Boer, D. Bakker, C, Vajda, I. Zuidgeest, M. (2017) PROM-Toolbox. https://www.zorginzicht.nl/kennisbank/Paginas/prom-toolbox.aspx.
8. Scholte M, Neeleman-van der Steen CW, van der Wees PJ, Nijhuis-van der Sanden MW, Braspenning J (2016) The Reasons behind the (Non)Use of Feedback Reports for Quality Improvement in Physical Therapy: A Mixed-Method Study. PLoS One 11 (8):e0161056.
9. Clement RC, Welander A, Stowell C, Cha TD, Chen JL, Davies M, Fairbank JC, Foley KT, Gehrchen M, Hagg O, Jacobs WC, Kahler R, Khan SN, Lieberman IH, Morisson B, Ohnmeiss DD, Peul WC, Shonnard NH, Smuck MW, Solberg TK, Stromqvist BH, Hooff ML, Wasan AD, Willems PC, Yeo W, Fritzell P (2015) A proposed set of metrics for standardized outcome reporting in the management of low back pain. Acta orthopaedica 86 (5):523-533.
10. Chiarotto A, Boers M, Deyo RA, Buchbinder R, Corbin TP, Costa LOP, Foster NE, Grotle M, Koes BW, Kovacs FM, Lin CC, Maher CG, Pearson AM, Peul WC, Schoene ML, Turk DC, van Tulder MW, Terwee CB, Ostelo RW (2018) Core outcome measurement instruments for clinical trials in nonspecific low back pain. Pain 159 (3):481-495.
11. Bombardier C (2000) Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. Spine 25 (24):3100-3103
12. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korff M, Waddell G (1998) Outcome measures for low back pain research. A proposal for standardized use. Spine 23 (18):2003-2013
13. Deyo RA, Dworkin SF, Amtmann D, Andersson G, Borenstein D, Carragee E, Carrino J, Chou R, Cook K, DeLitto A, Goertz C, Khalsa P, Loeser J, Mackey S, Panagis J, Rainville J, Tosteson T, Turk D, Von Korff M, Weiner DK (2014) Focus article: report of the NIH Task Force on Research Standards for Chronic Low Back Pain. European spine journal 23 (10):2028-2045.

14. Chiarotto A, Deyo RA, Terwee CB, Boers M, Buchbinder R, Corbin TP, Costa LOP, Foster NE, Grotle M, Koes BW, Kovacs FM, Lin C-WC, Maher CG, Pearson AM, Peul WC, Schoene ML, Turk DC, van Tulder MW, Ostelo RW (2015) Core outcome domains for clinical trials in non-specific low back pain. European Spine Journal 24 (6):1127-1142.

15. Rutten GM, Degen S, Hendriks EJ, Braspenning JC, Harting J, Oostendorp RA (2010) Adherence to Clinical Practice Guidelines for Low Back Pain in Physical Therapy: Do Patients Benefit? Physical Therapy 90 (8):1111-1122.

16. Hamilton AB, Yano EM (2017) The importance of symbolic and engaged participation in evidence-based quality improvement in a complex integrated healthcare system: response to "The science of stakeholder engagement in research". Translational behavioral medicine 7 (3):492-494.

17. Goodman MS, Sanders Thompson VL (2017) The science of stakeholder engagement in research: classification, implementation, and evaluation. Translational behavioral medicine 7 (3):486-491.

18. Burton AK, Balagué F, Cardon G, Eriksen HR, Henrotin Y, Lahad A, Leclerc A, Müller G, van der Beek AJ, On behalf of the CBWGoGfPiLBP (2006) Chapter 2 European guidelines for prevention in low back pain: November 2004. European Spine Journal 15 (Suppl 2):s136-s168.

19. Foster NE, Mullis R, Hill JC, Lewis M, Whitehurst DG, Doyle C, Konstantinou K, Main C, Somerville S, Sowden G, Wathall S, Young J, Hay EM (2014) Effect of stratified care for low back pain in family practice (IMPaCT Back): a prospective population-based sequential comparison. Annals of family medicine 12 (2):102-111.

20. Karran EL, McAuley JH, Traeger AC, Hillier SL, Grabherr L, Russek LN, Moseley GL (2017) Can screening instruments accurately determine poor outcome risk in adults with recent onset low back pain? A systematic review and meta-analysis. BMC medicine 15 (1):13.

21. Alrwaily M, Timko M, Schneider M, Stevans J, Bise C, Hariharan K, Delitto A (2016) Treatment-Based Classification System for Low Back Pain: Revision and Update. Phys Ther 96 (7):1057-1066.

22. Foster NE, Hill JC, O'Sullivan P, Hancock M (2013) Stratified models of care. Best practice & research Clinical rheumatology 27 (5):649-661.

23. Campbell SM, Braspenning J, Hutchinson A, Marshall MN (2003) Research methods used in developing and applying quality indicators in primary care. BMJ 326 (7393):816-819.

24. Fitch K, Steven J. Bernstein, Maria Dolores Aguilar, Bernard Burnand, Juan Ramon LaCalle, Pablo Lazaro, Mirjam van het Loo, Joseph McDonnell, Janneke Vader and James P. Kahan (2001) The RAND/UCLA Appropriateness Method User's Manual. RAND Corporation

25. Guideline International Network (G-I-N) database https://www.g-i-n.net.

26. Physiotherapy Evidence Database (PEDro). https://www.pedro.org.au/.

27. Staal J, Hendriks E, Heijmans M, Kiers H, Lutgers-Boomsma G, Rutten G (2013) KNGF-Guideline for low back painVerantwoording en toelichting. KNGF 21

28. Jones J, Hunter D (1995) Consensus methods for medical and health services research. BMJ: British Medical Journal 311 (7001):376

29. Cantrill J, Sibbald B, Buetow S (1996) The Delphi and nominal group techniques in health services research. International Journal of Pharmacy Practice 4 (2):67-74

30. Uscher-Pines L, Babin SM, Farrell CL, Hsieh YH, Moskal MD, Gaydos CA, Rothman RE (2010) Research priorities for syndromic surveillance systems response: consensus development using nominal group technique. Journal of public health management and practice : JPHMP 16 (6):529-534.

31. Tulder vMW, Custers JWH, Bie R. a. de, Hammelburg R, Kolhaar B, Kuijpers T (2010) nonspecific low back pain. CBO-guideline low back pain

32. van Tulder M, Becker A, Bekkering T, Breen A, del Real MT, Hutchinson A, Koes B, Laerum E, Malmivaara A (2006) Chapter 3. European guidelines for the management of acute nonspecific low back pain in primary care. European spine journal:S169-191.

33. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, Mannion AF, Reis S, Staal JB, Ursin H, Zanoli G (2006) Chapter 4. European guidelines for the management of chronic nonspecific low back pain. European spine journal: S192-300

34. Delitto A, George SZ, Dillen LV, Whitman JM, Sowa G, Shekelle P, Denninger TR, Godges JJ (2012) Low Back Pain. Journal of Orthopaedic & Sports Physical Therapy 42 (4):A1-A57.

35. Koes BW, van Tulder M, Lin CW, Macedo LG, McAuley J, Maher C (2010) An updated overview of clinical guidelines for the management of non-specific low back pain in primary care. European spine journal:2075-2094.

36. NHG-guideline for nonspecific low back pain (2005). https://www.nhg.org/standaarden/volledig/nhg-standaard-aspecifieke-lagerugpijn.

37. CBO guideline for nonspecific low back pain (2003). http://www.kwaliteitskoepel.nl/assets/structured-files/NOV/Aspecifiekelagerugklachten2003.pdf.

38. NICE (2016) Low back pain and sciatica in over 16s: assessment and management.

39. Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE, Hay EM (2008) A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis and rheumatism 59 (5):632-641.

40. Waddell G (2005) Subgroups within “nonspecific” low back pain. The Journal of rheumatology 32 (3):395-396

41. Westby MD, Klemm A, Li LC, Jones CA (2016) Emerging role of quality indicators in physical therapist practice and health service Delivery. Physical therapy 96 (1):90-100

42. Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, Williams JI (1995) The Quebec Back Pain Disability Scale. Measurement properties. Spine 20 (3):341-352

43. Calsbeek H, Maassen I, Braspenning JCC (november 2009) Publieke indicatoren eerstelijnszorg voor mensen met aspecifieke lage-rugpijn. Deel 8 uit de serie Voorhoedeproject Eerstelijns en Acute zorg – publieke indicatoren voor de eerstelijnszorg. Nijmegen: IQ healthcare

44. Van Dulmen SA, Schlief A, Oostendorp RA, Nijhuis-Van Der Sanden MW (december 2009) Einrapport project kwaliteitsindicatoren Fysiotherapie-3 (ProKwaF-3) Kwaliteit fysiotherapeutische verslaglegging. IQ healthcare

45. Fairbank JC, Couper J, Davies JB, O’Brien JP (1980) The Oswestry low back pain disability questionnaire. Physiotherapy 66 (8):271-273

46. Dutch Spine Surgery Registry vragenlijsten patientfeedback versie 3.0 2017 https://dica.nl/media/1255/DSSR%20Vragenlijsten%20V3.pdf.

47. van Hooff ML, van Loon J, van Limbeek J, de Kleuver M (2014) The Nijmegen decision tool for chronic low back pain. Development of a clinical decision tool for secondary or tertiary spine care specialists. PLoS One 9 (8):e104226.

48. Beurskens AJ, de Vet HC, Koke AJ, Lindeman E, van der Heijden GJ, Regtop W, Knipschild PG (1999) A patient-specific approach for measuring functional status in low back pain. Journal of manipulative and physiological therapeutics 22 (3):144-148

49. Bredero-Cohen AB, Meijer OG, Bruls VEJ, Luijnenburg-Kroes PE, Tulder vMW, Hendriks HJM, Martin L (April 2009) VvOCM Richtlijn aspecifieke lage-rugklachten https://vvocm.nl/LinkClick.aspx?fileticket=4yIvPoam1WU%3D&tabid=178&portalid=1&mid=660.

50. Numeric Pain Rating Scale (NPRS)

51. Hudak PL, Wright JG (2000) The characteristics of patient satisfaction measures. Spine 25 (24):3167-3177

52. Driehuis F, Van Dulmen SA, Nijhuis-Van Der Sanden MW, Van Der Wees PJ (2016) Eindrapportage Nelson Beattie: een vragenlijst naar patiënt ervaringen in de fysiotherapie IQ healthcare

53. Heeren J, Triemstra M, Boer D (2016) Ontwikkeling van de PREM Fysiotherapie; een vragenlijst over de ervaren kwaliteit van de fysiotherapeutische zorg. https://www.nivel.nl/sites/default/files/bestanden/Ontwikkeling-PREM-Fysiotherapie.pdf.

54. Vlaeyen JW, Kole-Snijders AM, Boeren RG, van Eek H (1995) Fear of movement/(re)injury in chronic low back pain and its relation to behavioral performance. Pain 62 (3):363-372

55. Heerkens Y, Hendriks H, de Graaf-Peters V (2011) KNGF guideline Fysiotherapeutische Verslaglegging Nederlands Tijdschrift voor Fysiotherapie 121:1–33

56. van Dulmen SA, van Der Wees PJ, Meerhoff GA, Maas MJM, Driehuis F, Bakker-Jacobs A, Cruijsberg J, Liefers J, Nijhuis-Van Der Sanden MW (04/2017) Eindrapport onderzoeksprogramma Masterplan Kwaliteit in Beweging (MKIB), Nijmegen: IQ healthcare.

57. Fagundes FR, de Melo do Espirito Santo C, de Luna Teixeira FM, Tonini TV, Cabral CM (2017) Effectiveness of the addition of therapeutic alliance with minimal intervention in the treatment of patients with chronic, nonspecific low back pain and low risk of involvement of psychosocial factors: a study protocol for a randomized controlled trial (TalkBack trial). Trials 18 (1):49.

58. Momsen AM, Jensen OK, Nielsen CV, Jensen C (2014) Multiple somatic symptoms in employees participating in a randomized controlled trial associated with sickness absence because of nonspecific low back pain. The spine journal : official journal of the North American Spine Society 14 (12):2868-2876

59. Wertli MM, Rasmussen-Barr E, Weiser S, Bachmann LM, Brunner F (2014) The role of fear avoidance beliefs as a prognostic factor for outcome in patients with nonspecific low back pain: a systematic review. The spine journal

60. George SZ, Beneciuk JM (2015) Psychological predictors of recovery from low back pain: a prospective study. BMC musculoskeletal disorders 16:49. doi:10.1186/s12891-015-0509-2

61. Janwantanakul P, Sitthipornvorakul E, Paksaichol A (2012) Risk factors for the onset of nonspecific low back pain in office workers: a systematic review of prospective cohort studies. Journal of manipulative and physiological therapeutics 35 (7):568-577.

62. Ramond-Roquin A, Bouton C, Begue C, Petit A, Roquelaure Y, Huez JF (2015) Psychosocial Risk Factors, Interventions, and Comorbidity in Patients with Non-Specific Low Back Pain in Primary Care: Need for Comprehensive and Patient-Centered Care. Frontiers in medicine 2:73.

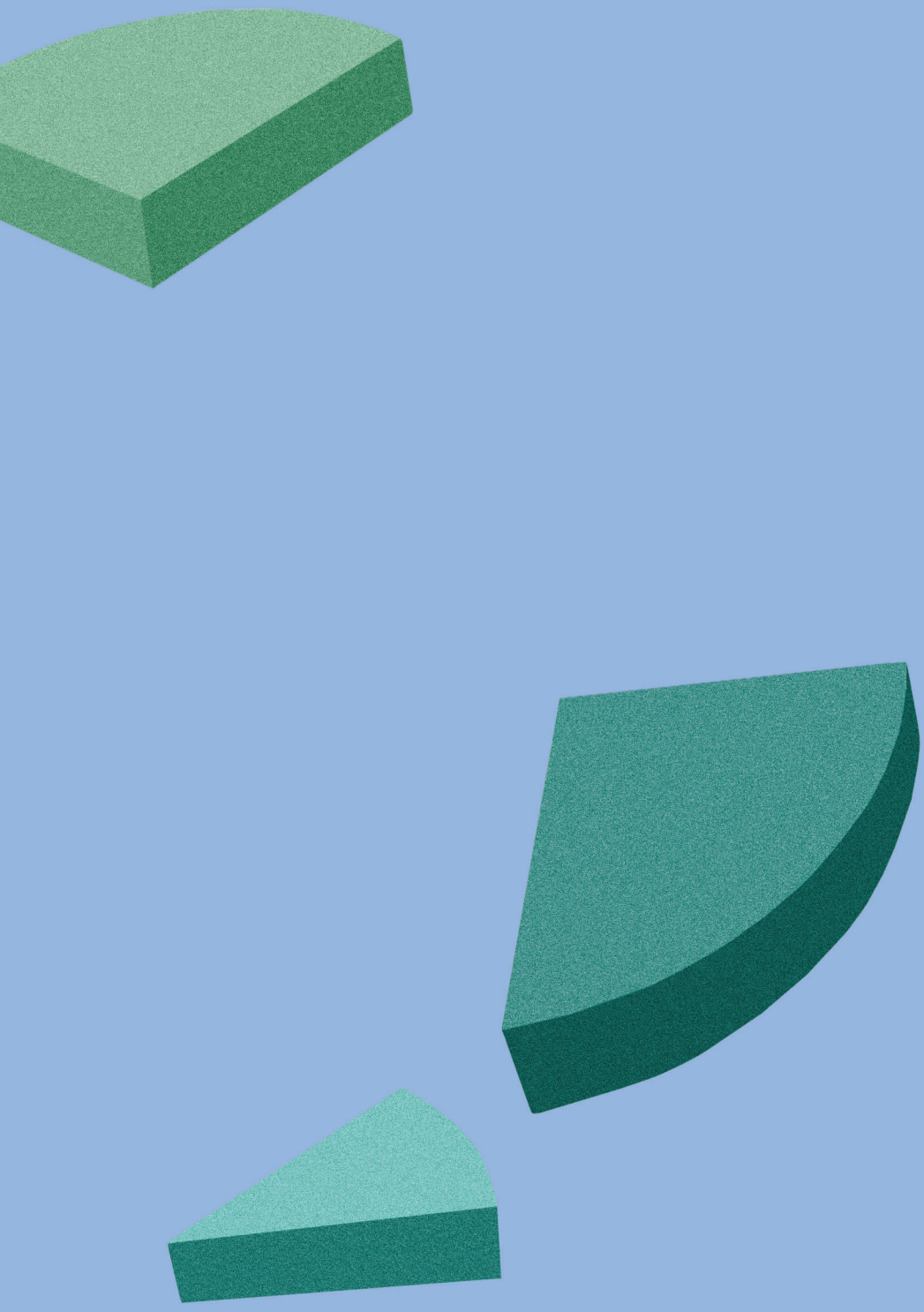
63. Pinheiro MB, Ferreira ML, Refshauge K, Ordonana JR, Machado GC, Prado LR, Maher CG, Ferreira PH (2015) Symptoms of Depression and Risk of New Episodes of Low Back Pain: A Systematic Review and Meta-Analysis. Arthritis care & research 67 (11):1591-1603.

64. Harms MC, Peers CE, Chase D (2010) Low back pain: what determines functional outcome at six months? An observational study. BMC musculoskeletal disorders 11:236.

65. Helmhout PH, Staal JB, Heymans MW, Harts CC, Hendriks EJ, de Bie RA (2010) Prognostic factors for perceived recovery or functional improvement in non-specific low back pain: secondary analyses of three randomized clinical trials. European spine journal:650-659.

66. Hill JC, Fritz JM (2011) Psychosocial influences on low back pain, disability, and response to treatment. Phys Ther 91 (5):712-721.

67. Ramond A, Bouton C, Richard I, Roquelaure Y, Baufreton C, Legrand E, Huez JF (2011) Psychosocial risk factors for chronic low back pain in primary care--a systematic review. Family practice 28 (1):12-21. doi:10.1093/fampra/cmq072



CHAPTER 3

Development of a standard set of outcome domains and proposed measures for chronic obstructive pulmonary disease in primary care physical therapy practice in the Netherlands

A modified RAND/UCLA appropriateness method

Arie C. Verburg
Simone A. van Dulmen
Henri Kiers
Jan H.L. Ypinga
Maria W.G. Nijhuis-van der Sanden
Philip J. van der Wees

International Journal of Chronic Obstructive Pulmonary Disease 2019:14 2649–2661
Digital object identifier: https://doi.org/10.2147/COPD.S219851

ABSTRACT

Background Standardization of measures in a common set opens the opportunity to learn from differences in treatment outcomes which can be used for improving quality of care. Furthermore, a standard set can provide the basis for development of quality indicators and is therefore useful for quality improvement and public reporting purposes. The aim of this study was to develop a standard set of outcome domains and proposed measures for patients with COPD in Dutch primary care physical therapy practice, including a proposal to stratify patients in subgroups.

Material and methods A consensus-driven modified RAND-UCLA Appropriateness method was conducted with relevant stakeholders (patients, physical therapists, researchers, policy makers and health insurers) in Dutch primary physical therapy care in eight steps: (1) literature search, (2) first online survey, (3) patient interviews, (4) expert meeting, resulting in a concept standard set and methods to identify subgroups (5) consensus meeting, (6) expert meeting (7) second online survey, and (8) final approval of an advisory board resulting of the approved standard set.

Results Five outcome domains were selected for COPD: physical capacity, muscle strength, physical activity, dyspnea and quality of life. A total of 21 measures were rated and discussed. Finally, eight measures were included, of which four mandatory measures: Characteristics of practices and physical therapists, Clinical COPD Questionnaire (CCQ)) for quality of life, Global Perceived Effect (GPE) for experience, 6-Minute Walk Test (6MWT) for physical capacity ; two conditional measures: Hand-Held Dynamometer (HHD) (with Microfet™) for Quadriceps strength, Medical Research Council Dyspnea (MRC) for monitoring dyspnea; and two exploratory measures: Accelerometry for physical activity, and the Assessment of Burden of COPD tool (ABC). To identify subgroups, a method described in the Dutch standard of care from the Lung Alliance was included.

Conclusion This study described the development of a standard set of outcome domains and proposed measures for patients with COPD in primary care physical therapy. Each measure was accepted for relevance and feasibility by the involved stakeholders. The set is currently used in daily practice and tested on validity and reliability in a pilot for the development of quality indicators.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is one of the common chronic health conditions of patients visiting a physical therapist in primary care practice in the Netherlands.[1] In 2017, a total of 35.227 patients with COPD were treated by primary care physical therapists.[1] Research showed that pulmonary rehabilitation can be beneficial for patients with COPD, for improving domains in health status including muscle function, cardiovascular function and reducing exacerbations.[2-4] Physical therapy is a key component of pulmonary rehabilitation for treatment of patients with COPD.[5-7] These outcome domains should be part of clinical practice. Standardization of outcome domains and proposed measures in a standard set opens the opportunity to learn from differences in treatment outcomes which can be used for improving quality of care.

Measures for outcomes measurement and quality improvement can focus on the patient's health status, e.g. scored with patient reported outcome measures (PROMs), patient reported experience measures (PREMs) and/or physical performance measures.[8-10] Interpretation of these measures over time can be used in the interaction between a patient and physical therapist (e.g. shared decision making, goal setting, and monitoring). Furthermore, physical therapists can learn from routine data collection to evaluate treatment effects and to compare differences in treatment effects between peers or other practices. Finally, a standard set of outcome domains and measures provides a basis for the development of quality indicators. Campbell et al (2003) defined a quality indicator as: "retrospectively measurable elements of practice performance for which there is evidence or consensus that it can be used to asses quality of care proved and hence change it".[11] Quality indicators consist of structure- (e.g. availability of a pulse oximetry device in the exercise area), process- (e.g. guideline adherence, shared decision making and goal setting) and outcome measures (e.g. quality of life or number of exacerbations).[11]

Successful implementation of a standard set of quality indicators in daily practice is challenging.[12] The first step in the development of quality indicators is the selection of a standard set of measures, and stakeholder engagement in reaching consensus is a key component for quality improvement initiatives.[13,14] This stresses the value of including all stakeholders in the development process.

The population of patients with COPD is heterogeneous with variation in the number of comorbidities, different levels of functional impairments, and as result differences in quality of life over time.[15,16] There is an increasing interest in stratification of heterogeneous patient groups to identify clustered characteristics within this group.[17-19] By identifying subgroups for patients with COPD, the outcomes of physical therapy treatment can be predicted more accurately. Identifying subgroups is therefore useful to validate the comparability between treatment outcomes in physical therapy practice.

Furthermore, stratified care enables the opportunity to provide personalized care by choosing relevant interventions for patients with COPD based on their characteristics.

The aim of this study was to develop a standard set of outcome domains and proposed measures for patients with COPD in Dutch primary care physical therapy practice. The final set of proposed measures should be accepted as having expected added value in clinical practice. The measures should enable comparing differences in treatment outcomes at the level of the individual patient, (monitoring and clinical decision making) and learning from differences in aggregated treatment outcomes of individual physical therapists or groups of physical therapists, for quality improvement purposes, and for public reporting. With including a proposal to stratify patients in subgroups, we aim to guarantee personal care and to enable comparing treatment outcomes between matched patients based on their characteristics.

METHODS

Design and setting

A RAND UCLA modified appropriateness method was conducted in primary care physical therapy practice in the Netherlands.[20] A mixed method approach was used between October 2016 and July 2017. Eight separate steps were performed to select eligible measures (table 1). The selection procedure was guided from the perspective of quality of care. This means that we focused on the content of care and the goals related to the treatment outcomes but also on the structure and process of care and patient experiences. Concerning the selection procedure for the outcome domains we first focused on the Dutch physical therapy guideline for COPD.[5] The guideline describes clinically relevant outcome domains using the International Classification of Functioning (ICF).[21,22] The ICF model is commonly used in physical therapy to determine outcome domains for patients based on their diagnosis, including functions (e.g. decreased exercise tolerance), activities (e.g. physical activity) and participation (e.g. quality of life).[22] A priori, our main focus was on the ICF outcome domains to select a standard set of proposed outcomes measurements including patient reported outcome measures (PROMs) and physical performance measures. However, because we hypothesized that structure- and process measures, including patient reported experience measures (PREMs) can also have an added value for evaluating quality of care, these were also included in the literature search. In order to stimulate stakeholder engagement an advisory board was formed with one representative of patients- i.e. the Lung Foundation Netherlands (Longfonds), representatives of physiotherapists: the Royal Dutch Society for Physical Therapy (KNGF), the Association for Quality in Physical Therapy (SKF), and two representatives of Dutch health insurance companies: CZ Health Insurance & The Friesland Insurance (DFZ). We followed the Dutch government statement that requires that quality indicators need to be accepted tripartite by

patient representatives, healthcare professionals and health insurers.[23] The board evaluated the process and was asked for their approval of the final standard set. The study protocol was approved by the Radboudumc Medical Ethical Committee (Registration # 2017-3154).

Table 1 Steps during the consensus-driven RAND/UCLA modified Delphi technique

Step	Participants	Goal	(Aimed) Results
1 Explorative review of the literature	Research team (AV, SD, RN, PW)	Identification of valid structure-, process-, and outcome (PROMs, PREMs, performance measures) measures for COPD that are described in scientific literature. Screening for methods to classify patients in subgroups.	Selection and analysis of all eligible measures on their validity and reliability. Selection methods to classify patients in subgroups.
2 First online survey round	22 out of 37 physical therapists specialized in COPD or scientific experience on COPD	Anonymous rating of measures with a 9-point Likert scale on relevance and feasibility.	The median score of every measure was calculated.
3 Expert committee	Five participants experts of step 2	Results of the online survey round 1 was interpreted and discussed with experts	First interpretation of the results of the first online survey presented and discussed
4 Interview with patients	Nine patients with COPD and treated by physical therapists last year	Patients views measurements in clinical practice	The patients perspectives on the use of measurement instruments which were presented during the introduction of the consensus meeting (step 5)
5 Consensus meeting	19 out of 37 participants of step 2 and members of the advisory board	Nominal group technique was used to discuss the draft set. Finally, the participants voted on feasibility and relevance.	Measures were included if 80% or higher voted yes for inclusion. Between 60 – 80 % yes were deferred and rerated in the second online survey and between 0 and 60 % yes-were excluded.
6 Expert committee and patient representatives	Participants of step 3 and the patient representative of the advisory board	The bottlenecks from step 5 will be discussed and searched for possible solutions	Alternatives for measures without consensus in step 5.
7 Second online survey round	23 out of 37 participants of the first online survey (step 2)	The participants rated the second draft set and if needed alternatives on a 9-point Likert scale	Measures were included when all participants rated a median of 7 or higher on the 9-point Likert scale.
8 Final approval of the advisory board	Advisory board (MJS, LV, HW, AW, AT and ML)	Finally, the advisory board was asked to accept the final set	Final acceptance of the standard set

Development process

The development process included an iterative process in eight consecutive steps: each step provided input for the following step.

Step 1 literature search

Potential measures for the standard set were searched in existing guidelines based on the outcome domains. We used a pragmatic explorative approach and did not aim at conducting a full systematic review of the literature. The databases of the Guidelines International Network (G-I-N) and Physical therapy Evidence Database (PEDro) were searched for guideline-based measures for patients with COPD.[24,25] We included all multi- and monodisciplinary Dutch and international clinical practice COPD guidelines for physical therapists, general practitioners, and medical specialists. Measures were also searched through websites of organizations developing clinical practice guidelines (see Appendix A). In the next phase, all eligible measures for outcome domains, process and structure measures relevant for patients with COPD were selected. Reasons for exclusion were: not familiar in the Netherlands, not developed and/or not useful for patients with COPD, not primarily recommended in guidelines, or not recommended for physical therapy primary care purposes. Per selected measure the following information was summarized: type of measure (process, structure, outcome); outcome domain (e.g. physical capacity or quality of life); whether it was a PROM, PREM or physical performance measure; content of the measure; time to complete the measure; the minimal clinically important difference (MCID); related measures; whether the measure was already translated in Dutch. We prioritized measures that were already used in daily practice in the Netherlands. In addition, we collected supporting literature about reliability , validity and responsiveness of the measure.

To search for methods to categorize patients with COPD in subgroups, the clinical guidelines identified in the literature search were used. Reference checking was performed to search for additional publications.

Step 2 first online survey

In total, 37 individuals were invited to participate in an online survey; including physical therapists specialized in COPD (n=25), senior researchers (n=3), policy makers (n=3) representative of a patient association (n=1), and representatives of regional networks of physical therapists specialized in COPD (n=5). Participants were recruited via the contact networks of the advisory board. Each participant needed to have at least 5 years' experience in treatment, research, or representing patients with COPD. The following two questions were scored for each measure selected from the first step: 1) Is this measure relevant to evaluate the quality of the physical therapy treatment for patients with COPD?; and 2) Is this measure feasible to score at the beginning and the end of the treatment episode for patients with COPD? The

selected measures and methods for classifying subgroups were scored using a 9-point Likert scale ranging from totally not relevant/feasible (0) to highly relevant/feasible (9).[20] As advised in the RAND/UCLA appropriateness method, the measures that scored a median of ≥7 on relevance and feasibility were initially accepted as measures having added value in daily practice.[20] The online survey was performed in LimeSurvey version 2.06.

Step 3 expert meeting

To interpret the data of the second step, we formed an expert group consisting of five expert physical therapists (EB, ET, CZ, ML, NP, and AH) who also participated in the online survey and had complementary expertise in treatment of patients with COPD, with a mean age of 50 years, and 50% was female. During a face-to-face meeting the experts reflected on the measures based on supporting literature regarding the validity, reliability and responsiveness of the proposed measures and their own experience. In addition, the experts were asked to interpret, discuss and if needed modify the methods to classify patients in subgroups that were found in the literature search.

Step 4 patient interviews

In each of nine purposefully selected physical therapy practices a patient with COPD was recruited. Purposeful sampling is a strategy in qualitative research to identify and select cases with rich information regarding the subject of interest and is highly appropriate for mixed methods studies.[26] Potential patients needed to be 18 years or older, diagnosed with COPD and treated by a physical therapist in the last year. The physical therapist gave the contact details of the patient to the researchers after the patient signed an informed consent form. Two researchers conducted semi-structured interviews (AV and JL) of 30 minutes by telephone using an interview guide (see appendix B). The patients were questioned about their perspectives on relevance and feasibility of outcome domains and proposed measures (PROMs, performance measures and PREMs), and if the measures are useful for the evaluation of physical therapy treatment. Furthermore, we asked patients about their views on relevance and feasibility of the structure, process and outcome measures. All interviews were audio-recorded, transcribed and analyzed using thematic analysis.[27] AV and JL independently analyzed the interviews and assigned codes within and between the interviews in quality data software ATLAS.ti 7.0. The assigned codes were compared and sort together in categories and themes.[27]

Step 5 consensus meeting

All 37 individuals of step 2 were invited for a three-hour face-to-face consensus meeting to select measures for the final standard set. The advisory board was invited to monitor the process. We used the Nominal Group Technique (NGT) to structure the meeting.[28,29] The NGT is a structured brainstorming process and opens the opportunity for all participants to contribute to the discussion.[30] First, the results of

step 1 to step 4 were presented and discussed. Then, the participants voted for inclusion of each initially selected measure and the classification in subgroups by raising their hand through a yes/no vote. The measure was included if ≥80% voted yes for inclus*ion.*[31] For the measures with a vote of yes between 80% and 60%, alternatives for the proposed measure were discussed with the expert group (step 6) and re-rated in a second online survey (step 7). Measures that scored ≤ 60% were excluded and not considered in the following steps. As part of the NGT, the group had the opportunity to suggest measures that were not initially selected after the literature search.[29] These newly suggested measures were rated and those that scored higher than 60% yes of the votes were discussed in the next step with the expert group.

Step 6 expert- and patient representative meeting

The measures for which no consensus (yes votes between 60-80%) was reached and the newly suggested measures (yes votes higher than 60%) in step 5 were discussed with the expert group, and separately with the Dutch Lung Foundation to obtain additional input from the patient's perspective. This was an iterative process through a face-to-face expert meeting, followed by consultation by telephone and group discussion via email.

The expert group was asked to suggest eligible alternatives that might be relevant and feasible in daily practice. Furthermore, the expert group developed a guided measurement protocol that provides an overview for physical therapists at what time points the measures need to be completed during the diagnosis and treatment process. Lastly, the expert group was asked to select case-mix variables for the standard set to identify patient characteristics and disease specific characteristics. When adjusting for these case-mix variables during analysis of treatment outcomes of patients with COPD, interpretation of the standard set is expected to be more accurate.

Step 7 Second online survey

The 37 participants of step 2 were invited for the second online survey. Alternative measures that were suggested by the expert group and the Dutch Lung Foundation were scored on a 9-point Likert scale. The measures were included if they scored a median of ≥7.[20]

Step 8 final approval of the advisory board

In the last step the standard set was presented to the advisory board and if they accepted the standard set as having added value, they were asked to sign an official approval document.

RESULTS

Step 1 literature search

After screening nine clinical guidelines [5,15,32-38] and additional literature on reliability and validity [39-45] , 45 measures were found and 21 measures were included. The reasons for exclusion of the 24 measures were: not familiar in the Netherlands, not developed and useful for patients with COPD, not primarily advised by guidelines, and not for physical therapy primary care purposes. Appendix C shows an overview of all measures. The 21 included measures consisted of one structure measure, two process measures, and 18 outcome measures, including 11 PROMs, two PREMs, three physical performance measures, and two other described outcome measures derived from a quality indicator. The following ICF outcome domains were selected: physical capacity, muscle function, dyspnea, physical activity and quality of life. Furthermore, we used elements of different clinical guidelines to propose a combined classification method for stratification of patients in subgroups based on the burden of disease, see appendix D.[5,32-35,37]

Step 2 first online survey

A total number of 22 out of 37 individuals accepted the invitation (response rate: 60%, mean age 46 years, 47% female), including physical therapists specialized in COPD (n=14), senior researchers (n=3), and representatives of regional networks of physical therapists specialized in COPD (n=5). After analyzing the results of the survey, 7 measures scored a median of ≥7 on relevance and feasibility, and also the proposal to stratify patients in subgroups based on the burden of disease (see appendix D) scored a median of ≥7. The other measures scored a median ≤6 for relevance and/or feasibility. Table 2 presents all rated measures.

Table 2 Results of step 2: first online survey and step 5: consensus meeting

Measures	**Relevance (step 2)**	**Feasibility (step 2)**	**Inclusion for the final standard set (step 5)**	
Structure measure	Median	Median	Yes	No
Characteristics of practices and physical therapists	8	8	85%**	15%
Process measures	Median	Median	Yes	No
History taking described in the EHR	5	7	0%	100%
Treatment plan described in the EHR	6	8	0%	100%
Quality of life measure for patients with high burden of disease*			92%**	8%
Outcome measures	Median	Median	Yes	No
Improvement in activities	7	7	0%	100%
Number of treatment sessions	3	8	7%	93%
Treatment costs*			57%	43%
Measure physical activity*			100%**	0%

Table 2 Continued

Measures	Relevance (step 2)	Feasibility (step 2)	Inclusion for the final standard set (step 5)	
Patient Reported Outcome Measures (PROMs)	Median	Median	Yes	No
Vragenlijst Fysieke Activiteit (VFA)	4	7	17%	83%
International Physical Activity Questionnaire (IPAQ)	7	8	0%	100%
Medical Research Council Dyspnea (MRC)	6,5	9	100%**	0
COPD Assessment Test (CAT)	6	7	0%	100%
Nijmegen Screenings instrument (NCSI)	6	6	0%	100%
Clinical COPD Questionnaire (CCQ)	8	9	100%**	0%
Chronic Respiratory (Disease) Questionnaire (CR(D)Q)	6	6	0%	100%
St George's respiratory questionnaire (SGRQ)	6	5	0%	100%
Quality of life for respiratory illness questionnaire (QoLRIQ)	5	5	0%	100%
Respiratory Illness Questionnaire-monitoring (RIQ-mon 10)	5	6	0%	100%
The Assessment of Burden of COPD index	7	8	0%	100%
The Assessment of Burden of COPD tool*			100%**	0%
Patient Reported Experience Measures (PREMs)	Median	Median	Yes	No
Global Perceived Effect (GPE)	6	7	100%	0%
PREM-P	7	4	0%	100%
Physical performance measures	Median	Median	Yes	No
6-Minute Walk Test (6 MWT)	8	9	100%**	0%
Incremental Shuttle Walk Test (ISWT)	6	5	0%	100%
Hand Held Dynamometrie (HHD)	7	8,5	0%	100%
Using the HDD (with a Microfet™) for quadriceps strength*			100%**	0%
Method to classify patients in subgroups based on the burden of disease	Appropriate-ness	Feasibility		
	Median	Median		
Classify patients in subgroups based on the burden of disease	7	7		

Notes *newly suggested during the consensus meeting in step 5, these measures will be discussed with the expert group in step 6 ** Final inclusion in the standard set after rating

Step 3 expert meeting

The most important statement that the expert group made was that treatment goals of patients with COPD are not only based on improving health condition but also on maintaining health status. They expressed that COPD is a progressive disease with fluctuations in health condition, which influences the expected outcomes of physical therapy treatment. When in further research the standard set is used for the development of quality indicators, caution is required on determining norm values on outcomes of the measures.

In addition, the expert groups suggested further details for the combined classification method in subgroups derived from the literature search. The details were presented at the consensus meeting in step 5.

Step 4 patient interviews

The average age of the nine patients (66% female) that were interviewed was 64.0 years (Standard Deviation [SD] 4.0) and GOLD stage ranged between II (n=1), III (n=3) and IV (n=5). The GOLD stage is a classification of airflow limitation in COPD ranges between I (mild) and IV (very severe).[15] After the data analysis, five main themes emerged: 1) questionnaires 2) patient satisfaction, 3) number of treatment sessions, 4) clinical practice and 5) quality of care. See appendix E for all themes, categories and codes. Patients stated that the readability of questionnaires was good. They indicated that they would spend a maximum of 10-15 minutes completing questionnaires. Some patients mentioned that patient satisfaction is an element of quality of care. According to the patients, the number of treatment sessions provided by the physical therapist is not a proxy for quality. Some patients stated that an important requirement for quality of treatments is that the practice facility needs to be adequate for doing exercises and should contain helpful equipment like a treadmill, home trainer or leg press. There were different opinions whether PROMs are relevant for measuring quality. The patients differed in their definition of high quality of care, for example defined by the treatment effect (maintaining health status), patient centeredness (communication, the physical therapist listens to me) or being coached by well-educated and specialized physical therapists.

Step 5 consensus meeting

In the consensus meeting, 19 individuals were present (response rate: 51%, mean age 43 years, and 63% female), including physical therapists specialized in COPD (n=10), senior researchers (n=3), policy makers (n=3), representatives of regional networks of physical therapists specialized in COPD (n=3), and the advisory board (n=5) to monitor the process. Their final votes for inclusion were for five initially selected measures, including: a structure measure with characteristics of practices and physical therapists, and the outcome measures Clinical COPD Questionnaire (CCQ) for quality of life, Medical Research Council Dyspnea (MRC) for dyspnea,

Global Perceived Effect (GPE) for experience, and the 6-Minute Walk Test (6 MWT) for physical capacity. All other measures were excluded by the participants.

Some individuals in the group suggested five alternatives for measures that were already derived from the literature search but not selected for the first survey: 1) Adding an additional quality of life questionnaire for patients with a high burden of disease (votes: 92% yes and 8% no); 2) Include treatment costs for value based healthcare purposes (votes: 57% yes and 43% no); 3) Monitoring physical activity in daily life, as an additional measure that requires further testing in the standard set (votes: 100% yes and 0% no); 4) The Assessment of Burden of COPD tool, also as an additional measure to evaluate whether the questionnaire is useful for evaluating physical therapy interventions (votes: 100% yes and 0% no); 5) The Hand-Held Dynamometer (HHD) (with a Microfet™) for monitoring quadriceps muscle strength (votes: 100% yes and 0% no). All suggested alternative measures were discussed in the next round, except suggestion 2.

During the meeting no consensus was reached about the inclusion of the suggested classification in subgroups from the literature. According to the participants, the presented classification method in subgroups was insufficient to determine the prognostic course of the patient group. The group agreed to discuss eligible alternatives with the expert group and patient representatives in the next step. See table 2 for all votes.

Step 6 expert- and patient representative meeting

In this round the expert group discussed a) alternatives to classify patients in subgroups, b) how to monitor physical activity in daily life, c) adding a general quality of life questionnaire, d) using HDD (with a Microfet™) for monitoring quadriceps muscle strength e) development of a guided treatment protocol, and finally, f) selecting case-mix variables:

(a) The expert group concluded after an iterative discussion that the described method for classifying subgroups can be replaced by a method that has already been used in the Netherlands, described in the multidisciplinary Dutch Care Standard of the Lung Alliance.[34] This guideline was also part for the combined classification method as described in step 1. The method describes three subgroups classified as light, moderate and high, based on the burden of disease.[34] The method is based on cut-off points of the MRC, CCQ, and number of exacerbations, lung function, and the body mass index (BMI) of the patient.

(b) The expert group agreed on monitoring physical activity as an additional measure that requires further testing in the standard set. The physical therapists

could choose whether they use a questionnaire, activity diary, accelerometer, or other activity trackers for measuring physical activity.

(c) The expert group advised that physical therapists could choose one of three eligible quality of life questionnaires as additional measure in the final standard set. The following questionnaires were suggested: Nijmegen Clinical Screening Instrument (NCSI), St George's Respiratory Questionnaire (SGRQ) and the COPD Assessment Test (CAT).

(d) The expert group agreed on the suggestion of the participants in step 5 to include the HDD (with a Microfet™) for monitoring quadriceps strength.

(e) The expert group was asked to establish a guided measurement protocol that provides an overview for physical therapists at which time points the measures need to be scored during the treatment. They advised to differentiate between mandatory measures for the total patient group, conditional measures depending on the treatment goals, and exploratory measures that require further testing. Mandatory measures included the selected structure measure with required characteristics of the practices and the physical therapists, and the outcome measures CCQ, GPE and 6MWT. The Dutch Lung Foundation preferred to measure the GPE every three months. Conditional measures, only relevant for specific treatment goals were using the HDD (with a Microfet™) for quadriceps muscle strength, and the MRC for dyspnea. Lastly, exploratory measures were included for monitoring physical activity and the Assessment of Burden of COPD tool.

(f) Finally, the expert group chose case-mix variables for a more accurate interpretation of the outcomes of the standard set. The case-mix variables included patient characteristics (age, gender, weight and length) and disease specific variables (lung values, smoke history, comorbidities, treatment goals and exacerbations). See table 4 for an explanation of the case-mix variables.

Step 7 Second online survey

In this step, 23 individuals (response rate: 64%, mean age 46, 44% female) completed the second online survey, including physical therapists specialized in COPD (n=14), senior researchers (n=3), a policy maker (n=1) and regional networks of physical therapists specialized in COPD (n=5). The alternative method to classify patients in subgroups and measurement protocol scored a median of ≥7 and was therefore included in the final standard set. The alternative suggestions for measuring physical activity and using quality of life questionnaires were scored with a median of 6. Nonetheless, based on narrative suggestions and discussion by the expert group, the accelerometer - when relevant for the treatment goal - can be useful for monitoring physical activity. Therefore we included the accelerometer as an

exploratory measure to monitor the level of physical activity of patients. The additional quality of life questionnaires were excluded. See table 3 for the final standard set.

Step 8 final approval of the advisory board

The advisory board accepted the final outcome set as shown in table 3. All stakeholders signed an official approval document.

Table 3 Final standard set

nr	Domain	Measure	Guided measurement protocol		
		A: mandatory for all patients with COPD	**Intake**	**Every 3 months**	**End**
1	Practice/physical therapist level	Characteristics of practices and physical therapists	Once a year		
2	Physical capacity	6-Minute Walk Test (6 MWT)	X	X	X
3	Quality of life	Clinical COPD Questionnaire (CCQ)	X	X	X[a]
4	Experience	Global Perceived Effect (GPE)		X	X
		B: Conditional measures			
5	Muscle strength	HDD (with a Microfet™) for quadriceps strength	X	X	X
6	Dyspnea	Medical Research Council Dyspnea (MRC)	X	X	X
		C: Exploratory measures			
7	Physical activity	Accelerometer (for physical activity in daily life)			
8	ABC-Tool	The Assessment of Burden of COPD tool			
		D: Classifying subgroups			
9	Classify in subgroups	Classify subgroups based on the Dutch care standard of the Lung Alliance.[34]	Once a year		

Note: [a] After ≥12 months the CCQ needs only to be measured once a year.

Table 4 Case Mix Variables

Variable	Description
Age	Date of birth
Gender	Gender at birth
Weight	Weight in kg
Length	Length in cm
Post FEV1	In mL
Smoke history	1. Currently smoking 2. if yes, note how many cigarettes a week 3. Did smoked 4. Never smoked
Comorbidities	1. Cardiac disorders 2. Vascular disorders 3. Disorders of bones, muscles or the skin (e.g. contractures, osteoarthritis) 4. Psychosocial disorders (e.g. depression, addictions) 5. Endocrine and metabolic disorders, generalized infections, poisoning (osteoporosis, diabetes mellitus) 6. Other
Most important treatment goals Treatment goal	1. Dyspnea 2. Exercise capacity 3. Physical activity 4. Muscle strength 5. Self-management
Number of exacerbations last year	1. 0 2. 1 3. 2 4. 4 or more

DISCUSSION

In this consensus study a standard set was developed with five outcome domains: physical capacity, muscle strength, physical activity, dyspnea, quality of life, and eight proposed measures were selected for patients with COPD treated in primary care physical therapy practice. The standard set consists of four mandatory measures for all patients with COPD, including one structure measure; characteristics of practices and physical therapists, and three outcomes measures; the 6-Minute Walk Test (6 MWT) for physical capacity, the Clinical COPD Questionnaire (CCQ) for quality of life, and the Global Perceived Effect (GPE) for experience. Two conditional measures are included depending on the treatment goal: using the HDD (with a Microfet™) for quadriceps muscle strength, and Medical Research Council Dyspnea (MRC) for monitoring dyspnea. Two exploratory measures are included: use of an accelerometer for monitoring physical activity in daily life, and The Assessment of Burden of COPD tool (ABC). For identifying subgroups based on the burden of disease, the method of the Dutch care standard of the Lung Alliance is included. Finally, case mix variables were selected for a more accurate interpretation of the outcomes in the standard set.

The standard set was accepted as having expected added value in clinical practice and is therefore deemed useful for the interaction between a patient and a physical therapist. Furthermore, comparison of outcomes of the standard set between physical therapists on individual and group level opens the opportunity to learn from routine data collection, and finally the standard set provides a basis for development of quality indicators. To our knowledge, this is the first study that describes a standard set for patients with COPD for these specific goals including development of quality indicators in primary care physical therapy.

The recent study of Souto-Miranda et al, described domains for measures in a core outcome set based on stakeholders perspectives (patients, informal care providers and health professionals).[46] Important described needs (e.g. improving exercise tolerance and reducing dyspnea) by the stakeholders are in line with the measures in the standard set (e.g. 6MWT and MRC)[46]. However, there are also differences between the study of Souto-Miranda et al and this study. The goal of our standard set is developing quality indicators and enhance quality improvement initiatives in clinical practice. The described goals of the study of Souto-Miranda et al are to inform on a core set that generates consistency among clinical trials and decrease risk of bias in research studies by standardizing outcomes.

The expert group stated in step 3 that patients with COPD are a heterogeneous patient group with more or less comorbidities and exacerbations that cannot always be influenced by physical therapists. They expressed that for that reason caution is required with interpreting outcomes between physical therapists and practices. We therefore included case-mix variables and a stratification tool to allocate patients in more homogeneous subgroups. Identifying these subgroups opens the opportunity to compare and predict outcomes more accurate for the same patient population. Using the standard set combined with stratification in subgroups and case-mix variables, opens the opportunity for physical therapists and practices to use quality indicators as a learning tool for quality improvement initiatives by comparing outcomes between their peers.[19,47] Nonetheless, when comparing and interpreting outcomes of physiotherapists and practices the fluctuating health condition of patients with COPD needs to be included.[48]

As described in the introduction, stakeholder engagement is highly important in the development of quality indicators. Therefore, we were pleased that all stakeholders were included in the consensus rounds. Zorgverzekeraars Nederland (ZN)[49], which is the umbrella organization of the ten health insurance compagnies in the Netherlands, accepted the outcomes of the final standard set and agreed to use the standard set in the development of quality indicators.

Our study has several limitations. One limitation of our study is that the consensus rounds were conducted in the Netherlands and focused on the Dutch healthcare system. In addition, we selected outcome domains based on an Dutch guideline for physical therapists,

and preferred proposed measures that were already used in the Netherlands. The generalizability for international use of the standard set of outcome domains and proposed measures may therefore be limited. However, the selection of measures was based on a literature review of international clinical practice guidelines. This is in line with other studies developing quality indicators e.g. the study of Westby 2018 et al.[50] Still the context of each country needs to be taken into account, and we think the focused method of our study was helpful to encourage successful implementation in Dutch physical therapy practice.[51]

The COMET initiative provides guidance for the selection of outcome domains and outcome measures in developing core outcome sets.[52] The COMET initiative is a valuable and important initiative to develop and inform on core outcome sets (COS) for clinical trial purposes and clinical auditing. We chose to use the RAND/UCLA appropriateness method, which is widely used and provides a manual for synthesizing expert opinion and evidence for the development of quality indicators.[11,47,53] The steps described by COMET are to a large extent similar with the RAND/UCLA appropriateness method, including literature review, Delphi procedure and face-to-face consensus meetings; the purpose though is different. In addition, the COMET handbook version 1.0 was published after conduction of our study.[52]

After this consensus study, the standard set was implemented for a pilot test in the Netherlands for the development of quality indicators. In January 2018 a large pilot was launched where over 250 physical therapists started using the standard set for collecting data of approximately 4000 patients with COPD treated in primary care. In the Netherlands, many software systems for electronic health records (EHR) are used in primary care. The EHRs must be connected to a secure database. Also, the data collection of all EHRs needs to be standardized with the standard set; otherwise comparison of the outcomes can be invalid. An important requirement for successful data collection of the standard set is that the infrastructure is adequate.

This study presents a standard set of outcome domains and proposed measures for patients with COPD in primary care physical therapy; each measure is accepted for relevance and feasibility by stakeholders. The standard set is a promising basis for development of quality indicators in primary care physical therapy practice.

We acknowledge the following stakeholders, experts and colleagues for their participation in our study: Expert committee Emmylou Beekman (EB), Ellen Toet (ET), Cor Zagers (CZ), Monica van der Lans (ML) and Nic van Paassen (NP). Advisory board The Royal Dutch Society for Physical Therapy (KNGF): Annemarie Trompert (AT). The Association for Quality in Physical Therapy (SKF): Marije de Leur (ML). The Lung Foundation Netherlands (Longfonds): Marie José Schrasser (MS). Health insurance company CZ Group: Hans Wattel (HW). Health insurance company de Friesland: Alida Wolters (AW). Research members Joan Luites (JL) and Juliette Cruijsberg (JC).

REFERENCES

1. Vektis https://www.zorgprismapubliek.nl/producten/zorg-dichtbij/fysiotherapie/row-5/voor-welke-diagnoses-binnen-de-basisverzekering-komen-mensen-het-meest-bij-de-fysiotherapeut/. Accessed.
2. Spruit MA, Singh SJ, Garvey C, et al. An official American Thoracic Society/European Respiratory Society statement: key concepts and advances in pulmonary rehabilitation. *American journal of respiratory and critical care medicine.* 2013;188(8):e13-64.
3. McCarthy B, Casey D, Devane D, Murphy K, Murphy E, Lacasse Y. Pulmonary rehabilitation for chronic obstructive pulmonary disease. *The Cochrane database of systematic reviews.* 2015(2):Cd003793.
4. Puhan MA, Gimeno-Santos E, Scharplatz M, Troosters T, Walters EH, Steurer J. Pulmonary rehabilitation following exacerbations of chronic obstructive pulmonary disease. *The Cochrane database of systematic reviews.* 2011(10):Cd005305.
5. Gosselink R, Langer D, Burtin C, et al. KNGF-Guideline for physical therapy in patients with chronic obstructive pulmonary diseasehttps://www.fysionet-evidencebased.nl/images/pdfs/richtlijnen/copd_2008/copd_verantwoording_en_toelichting.pdf. 2008.
6. Beekman E, Mesters I, Hendriks E, et al. Exacerbations in patients with chronic obstructive pulmonary disease receiving physical therapy: A cohort-nested randomised controlled trial. *BMC pulmonary medicine.* 2014;14:71.
7. Leeden M, Staal JB, Beekman E, et al. Development of a framework to describe goals and content of exercise interventions in physical therapy: A mixed method approach including a systematic review. *Physical Therapy Reviews.* 2014;19:1-14.
8. Meerhoff GA, van Dulmen SA, Maas MJM, Heijblom K, Nijhuis-van der Sanden MWG, Van der Wees PJ. Development and Evaluation of an Implementation Strategy for Collecting Data in a National Registry and the Use of Patient-Reported Outcome Measures in Physical Therapist Practices: Quality Improvement Study. *Physical Therapy.* 2017;97(8):837-851.
9. Van Der Wees PJ, Nijhuis-Van Der Sanden MW, Ayanian JZ, Black N, Westert GP, Schneider EC. Integrating the use of patient-reported outcomes for both clinical practice and performance measurement: views of experts from 3 countries. *The Milbank quarterly.* 2014;92(4):754-775.
10. Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use.* Oxford University Press, USA; 2015.
11. Campbell SM, Braspenning J, Hutchinson A, Marshall MN. Research methods used in developing and applying quality indicators in primary care. *BMJ.* 2003;326(7393):816-819.
12. Brand C, Landgren F, Hutchinson A, Jones C, MacGregor L, Campbell D. Clinical practice guidelines: barriers to durability after effective early implementation. *Internal Medicine Journal.* 2005;35(3):162-169.
13. Hamilton AB, Yano EM. The importance of symbolic and engaged participation in evidence-based quality improvement in a complex integrated healthcare system: response to "The science of stakeholder engagement in research". *Translational behavioral medicine.* 2017;7(3):492-494.
14. Goodman MS, Sanders Thompson VL. The science of stakeholder engagement in research: classification, implementation, and evaluation. *Translational behavioral medicine.* 2017;7(3):486-491.

15. Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global Strategy for the Diagnosis, Management and Prevention of Chronic Obstructive Lung Disease 2017 Report: GOLD Executive Summary. *Respirology (Carlton, Vic).* 2017;22(3):575-601.

16. Licher S, Heshmatollah A, van der Willik KD, et al. Lifetime risk and multimorbidity of non-communicable diseases and disease-free life expectancy in the general population: A population-based cohort study. *PLOS Medicine.* 2019;16(2):e1002741.

17. Fritz JM. Clinical Prediction Rules in Physical Therapy: Coming of Age? *Journal of Orthopaedic & Sports Physical Therapy.* 2009;39(3):159-161.

18. Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. *BMJ : British Medical Journal.* 2013;346:e5595.

19. Hingorani AD, Windt DAvd, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *BMJ : British Medical Journal.* 2013;346:e5793.

20. Fitch K, Steven J. Bernstein, Maria Dolores Aguilar, Bernard Burnand, Juan Ramon LaCalle, Pablo Lazaro, Mirjam van het Loo, Joseph McDonnell, Janneke Vader and James P. Kahan. The RAND/UCLA Appropriateness Method User's Manual. *RAND Corporation.* 2001.

21. Escorpizo R, Stucki G, Cieza A, Davis K, Stumbo T, Riddle DL. Creating an interface between the International Classification of Functioning, Disability and Health and physical therapist practice. *Phys Ther.* 2010;90(7):1053-1063.

22. The ICF: An Overview available at: https://www.wcpt.org/sites/wcpt.org/files/files/GH-ICF_overview_FINAL_for_WHO.pdf.

23. Zorginstituut Nederland Toetsingskader kwaliteitsstandaarden, informatiestandaarden en meetinstrumenten 2015 – herziene versie juni 2018 available at: https://www.zorginstituutnederland.nl/publicaties/publicatie/2018/06/28/toetsingskader-kwaliteitsstandaarden-informatiestandaarden-en-meetinstrumenten-2015.

24. Guideline International Network (G-I-N) database https://www.g-i-n.net. Accessed 01 Dec. 2016.

25. Physiotherapy Evidence Database (PEDro). https://www.pedro.org.au/. Accessed 01 Dec. 2016.

26. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and policy in mental health.* 2015;42(5):533-544.

27. Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. *BMJ (Clinical research ed).* 2000;320(7227):114-116.

28. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ: British Medical Journal.* 1995;311(7001):376.

29. Cantrill J, Sibbald B, Buetow S. The Delphi and nominal group techniques in health services research. *International Journal of Pharmacy Practice.* 1996;4(2):67-74.

30. Uscher-Pines L, Babin SM, Farrell CL, et al. Research priorities for syndromic surveillance systems response: consensus development using nominal group technique. *Journal of public health management and practice : JPHMP.* 2010;16(6):529-534.

31. Craft Ralph C. The Pareto principle in organizational decision making. *Management Decision.* 2002;40(8):729-733.

32. Geijer RM, Tuut MK, in't Veen JC, Broekhuizen BD, Chavannes NH, Smeele IJ. [The NHG guidelines 'Adult asthma' and 'COPD']. *Nederlands tijdschrift voor geneeskunde.* 2015;159:A9076.

33. Dekhuijzen PNR, Geijer RM, Grol MH, Kerstjens HAM, Smeele IJM, Weijermans JW. Richtlijn Diagnostiek en behandeling van COPD: Actualisatie 2010. *Utrecht: Kwaliteitsinstituut voor de Gezondheidszorg CBO.* 2010.

34. Salome PL, in 't Veen JC, Euving E, et al. Dutch Lung Alliance (LAN) COPD Zorgstandaard 2016.

35. Koolen EH, van der Wees PJ, Westert GP, Dekhuijzen R, Heijdra YF, van 't Hul AJ. The COPDnet integrated care model. *International journal of chronic obstructive pulmonary disease.* 2018;13:2225-2235.

36. VBC. Vereniging Bewegingsleer Cesar (VBC) Richtlijn Cesar en COPD. 2004.

37. NICE. National Institute for Health and Care Excellence (NICE) Guideline, chronic obstructive pulmonary disease in over 16s: diagnosis and management. 2010.

38. Yang I, Dabscheck E, George J, et al. The COPD-X Plan: Australian and New Zealand guidelines for the management of chronic obstructive pulmonary disease 2016. *Thoracic Society of Australia and New Zealand and Lung Foundation Australia Available online at: http://copdx org au (accessed July 2015).* 2016.

39. Van Dulmen SA, Schlief A, Oostendorp RA, Nijhuis-Van Der Sanden MW. Einrapport project kwaliteitsindicatoren Fysiotherapie-3 (ProKwaF-3) Kwaliteit fysiotherapeutische verslaglegging. *IQ healthcare.* december 2009.

40. Driehuis F, Van Dulmen SA, Nijhuis-Van Der Sanden MW, Van Der Wees PJ. Eindrapportage Nelson Beattie: een vragenlijst naar patiënt ervaringen in de fysiotherapie *IQ healthcare* 2016.

41. Heeren J, Triemstra M, Boer D. Ontwikkeling van de PREM Fysiotherapie; een vragenlijst over de ervaren kwaliteit van de fysiotherapeutische zorg. https://www.nivel.nl/sites/default/files/bestanden/Ontwikkeling-PREM-Fysiotherapie.pdf. Published 2016 Accessed 01 Dec. 2016.

42. Calsbeek H, Maassen I, Braspenning JCC. Publieke indicatoren eerstelijnszorg voor mensen met aspecifieke lage-rugpijn. Deel 8 uit de serie Voorhoedeproject Eerstelijns en Acute zorg – publieke indicatoren voor de eerstelijnszorg. . *Nijmegen: IQ healthcare.* november 2009.

43. Heerkens Y, Hendriks H, de Graaf-Peters V. KNGF guideline Fysiotherapeutische Verslaglegging Nederlands Tijdschrift voor Fysiotherapie 121:1–33 2011.

44. Scholte M, Neeleman-van der Steen CW, van der Wees PJ, Nijhuis-van der Sanden MW, Braspenning J. The Reasons behind the (Non)Use of Feedback Reports for Quality Improvement in Physical Therapy: A Mixed-Method Study. *PLoS One.* 2016;11(8):e0161056.

45. van Dulmen SA, van Der Wees PJ, Meerhoff GA, et al. Eindrapport onderzoeksprogramma Masterplan Kwaliteit in Beweging (MKIB), Nijmegen: IQ healthcare. 04/2017.

46. Souto-Miranda S, Marques A. Triangulated perspectives on outcomes of pulmonary rehabilitation in patients with COPD: a qualitative study to inform a core outcome set. *Clinical rehabilitation.* 2018:269215518821405.

47. Westby MD, Klemm A, Li LC, Jones CA. Emerging role of quality indicators in physical therapist practice and health service Delivery. *Physical therapy.* 2016;96(1):90-100.

48. Rodriguez-Roisin R. Toward a consensus definition for COPD exacerbations. *Chest.* 2000;117(5 Suppl 2):398s-401s.

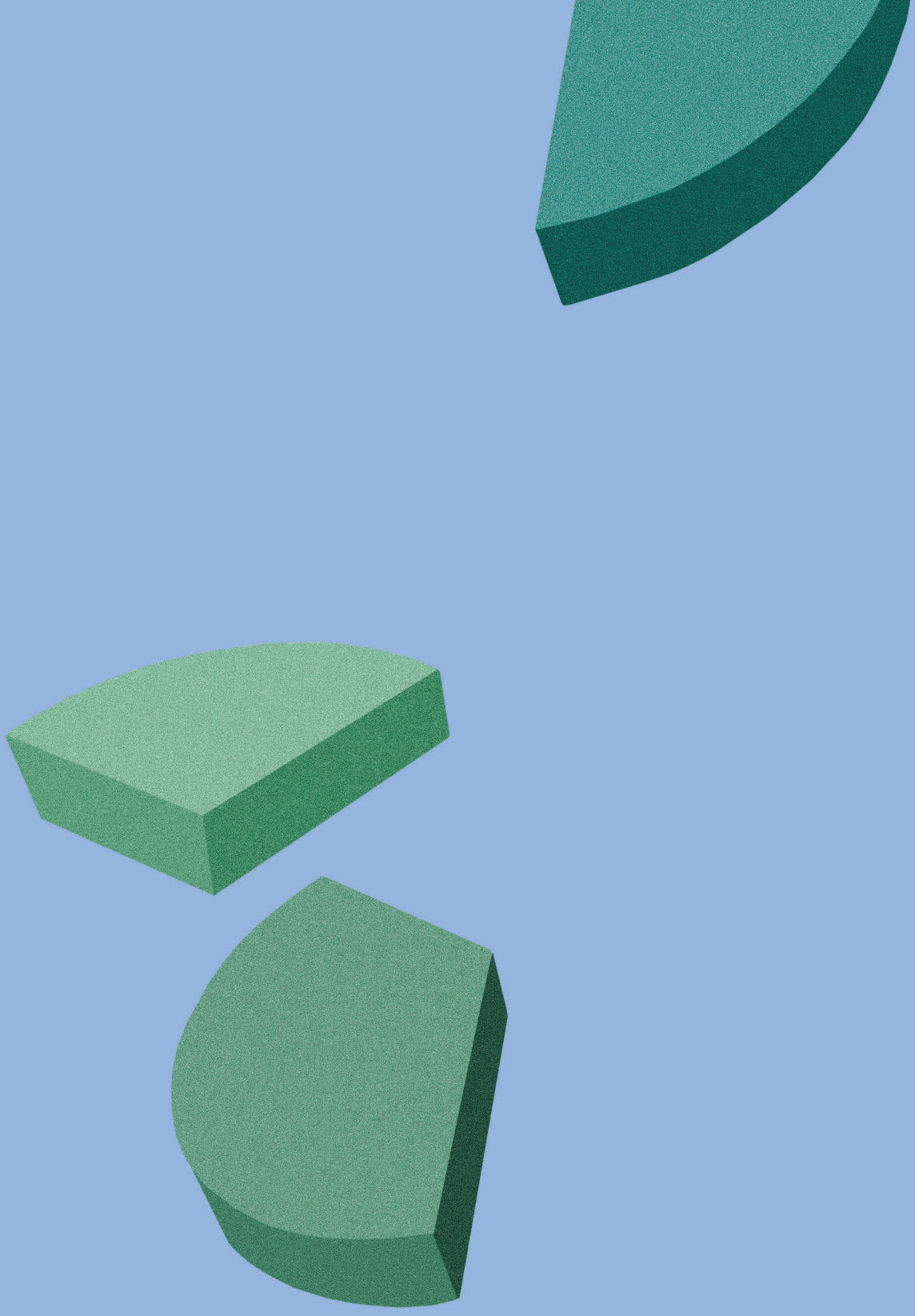
49. Zorgverzekeraars Nederland https://www.zn.nl/1483931648/About-ZN. Accessed.

50. Westby MD, Marshall DA, Jones CA. Development of quality indicators for hip and knee arthroplasty rehabilitation. *Osteoarthritis and cartilage.* 2018;26(3):370-382.

51. Marshall MN, Shekelle PG, McGlynn EA, Campbell S, Brook RH, Roland MO. Can health care quality indicators be transferred between countries? *Quality & safety in health care.* 2003;12(1):8-12.

52. Williamson PR, Altman DG, Bagley H, et al. The COMET Handbook: version 1.0. *Trials.* 2017;18(3):280.

53. Fitch KSJ, Bernstein. Maria, Dolores Aguilar. Bernard, Burnand. Juan Ramon, LaCalle. Pablo, Lazaro. Mirjam, van het Loo. Joseph, McDonnell. Janneke, Vader. and James P, Kahan. The RAND/UCLA Appropriateness Method User's Manual. Santa Monica, CA:. *RAND Corporation.* 2001.



CHAPTER 4

Patient-reported outcome-based quality indicators in Dutch primary care physical therapy for patients with nonspecific low back pain

A cohort study

Arie C. Verburg
Simone A. van Dulmen
Henri Kiers
Maria W.G. Nijhuis-van der Sanden
Philip J. van der Wees

Physical Therapy, Volume 101, Issue 8, August 2021
Digital object identifier: https://doi.org/10.1093/ptj/pzab118

ABSTRACT

Objective. The purpose of this study was to define and select a core set of outcome-based quality indicators, accepted by stakeholders on usability and perceived added value as a quality improvement tool, and to formulate recommendations for the next implementation step.

Methods. In phase 1, we defined 15 potential quality indicators for patient-reported outcome measures (PROMs) and associated domains, namely the Numeric Pain Rating Scale (NPRS) for pain intensity, the Patient Specific Functioning Scale (PSFS) for physical activity, the Quebec Back Pain Disability Scale (QBPDS) for physical functioning, and the Global Perceived Effect - Dutch Version (GPE-DV) for perceived effect. We described their comparability and discriminatory characteristics using cohort data. In phase 2, a core set of quality indicators was selected based on consensus among stakeholders in focus group meetings.

Results. In total, 65,815 completed treatment episodes for patients with nonspecific low back pain (NSLBP) were provided by 1009 physical therapists from 219 physical therapist practices. The discriminability between physical therapists of all potential 15 quality indicators was adequate with intraclass correlation coefficients between 0.08 and 0.30. Stakeholders selected a final core set of 6 quality indicators: 2 process indicators (the routine measurement of NPRS and the PSFS) and 4 outcome indicators (pretreatment and posttreatment change scores for the NPRS, PSFS, QBPDS, and the minimal clinically important difference of the GPE-DV).

Conclusion. This study described and selected a core set of outcome-based quality indicators for physical therapy in patients with NSLBP. The set was accepted by stakeholders for having added value for daily practice in physical therapy primary care and was found useful for quality improvement initiatives. Further studies need to focus on improvement of using the core set of outcome-based quality indicators as a quality monitoring and evaluation instrument.

INTRODUCTION

Quality indicators based on health outcomes can be used as a implementation tool to improve quality of care by monitoring and evaluating treatment trajectories with patient reported outcomes (PROMs).(1-4) Quality indicators are measurable items to monitor healthcare, referring to structure, processes or outcomes of care.(5) An important requirement is that these outcome based quality indicators are well developed, described and have added value in improving quality of care.(3, 6) Initial development of quality indicators is usually based on a theoretical perspective and consensus between stakeholders, while routinely collected clinical data should be included in the ecological validation process.(5) Currently no core set of outcome-based quality indicators is developed for patients with nonspecific low back pain (NSLB) in primary care physical therapy based on such a comprehensive process.

In a previous study we developed a standard set of outcome domains and patient reported outcome measures (PROMs) for patients with NSLBP in primary care physical therapist practice.(7) PROMs are questionnaires or single item scales to assess PROs in relevant outcome domains, such as pain, physical functioning, or perceived treatment effect.(5, 8) The standard set with PROMs is the basis for our next step in the development and implementation of outcome-based quality indicators, which preferably can be used for large patient groups and are able to show changes in clinical practice over time.(3, 7)

Various standards for development and selection of quality indicators exist.(5, 9) To develop quality indicators, firstly the comparability of the outcomes between physical therapists or practices needs to be adequate. When considering patient characteristics that may influence the outcome, but are not under the control of the physical therapist or practice, case-mix adjustment and stratification in subgroups can be used to improve the population comparability.(5, 9, 10) Secondly, in order to drive quality improvement, the quality indicator should be able to discriminate between physical therapists or practices based on the outcomes measures.(5, 9) Finally, quality indicators need to be accepted by stakeholders (i.e. physical therapists, patients and health insurers) based on the perceived added value and usability for quality improvement in daily practice. We used the Netherlands as case study for the development of outcome-based quality indicators.

(a) For development of outcome-based quality indicators for patients with NSLBP in Dutch physical therapy primary care practices, the aims of this study are:

(b) To define potential outcome-based quality indicators based on the previous selected standard set of outcome domains and selected PROMs, and describe their comparability and discriminability,

To select a core set from potential outcome-based quality indicators, accepted by stakeholders on usability and perceived added value as quality improvement tools.

METHODS

Design

In the current study we intended to develop quality indicators in consecutive steps, using a mixed methods approach in a sequential explanatory design. In phase 1 (October 2017-September 2019), we defined potential quality indicators and estimated their comparability and discriminatory characteristics using prospectively collected patient outcomes in a convenience cohort.

In phase 2 (October-December 2019), the outcomes were presented to stakeholders in a qualitative approach to explore the usability by interpreting their comparability, discriminability and perceived added value, in order to select a core set of quality indicators.

Setting

In total, 1,009 Dutch physical therapists working in 219 primary care practices collected outcomes of treatment trajectories of patients with NSLBP. All graduated, licensed Dutch physical therapists that treated patients with NSLBP could participate in the study, and they needed to be able to provide data to one of the three data registries, as specified in the subheading data collection. They were recruited using convenience sampling via stakeholder organizations (professional physical therapy associations and health insurers) in Dutch primary physical therapy care and participated voluntarily in the project. All procedures were conducted according to the Declaration of Helsinki. The study protocol was approved by the Medical Ethical Committee of Radboud university medical center (Registration # 2019-5455). Funding support for this study was provided by Dutch health insurers CZ & the Friesland. The sponsors had no influence on the process or outcomes of the paper.

Data collection

Treatment outcomes were anonymously collected through electronical health records (EHRs). All data were collected via three databases, the national data registry (LDK) of the Association for Quality in Physical Therapy (SKF), the national data registry (LDF) of the Royal Dutch Society for Physical Therapy (KNGF), and the database of Spot On Medics (SOM) - which is one of the EHR vendors. Participating physical therapists were instructed to seek informed consent of each included patient to use the data from the EHR for research and quality improvement.

PROMs

A standard set of PROMs and associated outcome domains was previously selected as relevant for clinical practice and as basis for the development of quality indicators.(7) This standard set included the Numeric Pain Rating Scale (NPRS) for measuring pain intensity, the Patient Specific Functional Scale (PSFS) for measuring physical activity, and the Quebec Back Pain Disability Scale (QBPDS) for measuring physical function,

which were collected pretreatment and posttreatment to monitor changes in outcomes over time. The Global Perceived Effect - Dutch Version (GPE-DV) for measuring perceived effect was completed by patients at the end of the treatment episode to evaluate treatment effect. See Supplementary Appendix A for an overview of each measure, measurement protocol and a comprehensive description of all measures.

Case-mix Adjustment and Stratification

For each patient we collected the following characteristics for case-mix adjustment: age, sex, and chronicity (expressed in the duration of the complaints before treatment). The Start Back Tool (SBT) was administered at the beginning of the treatment episode, and used to stratify patients in a low, medium or high-risk profile.(11)

Inclusion and Exclusion Criteria

Participating physical therapists and practices included their patients aged ≥18 years with NSLBP. For each patient we aimed to collect the PROMs in the standard set as described in the data collection. The patients received usual care from the participating physical therapists according to the Dutch clinical practice guidelines for NSLBP, meaning that we only measured outcomes of the treatment, the physical therapists individually decided which treatment was needed for their patients. (12) Each physical therapist or practice should at least include >5 patients during the data collection period to participate in this study.

We included only patients with a closed treatment episode. An episode was considered as closed when the physical therapist closed the episode in the EHR or if eight weeks had been passed after the last visit. The same patient can have more treatment episodes in a year, which in this study were handled as separate unique episodes."

Phase 1: Defining and describing the comparability and discriminability of the quality indicators

Defining potential quality indicators

Potential quality indicators were defined using national and international standards.(5, 9, 10, 13) Quality indicators can be quantified and expressed as a percentage using a denominator and a numerator.(3) The denominator usually describes the number of persons in the target group for which the quality indicator is applicable. In the numerator, the number of 'correct' scores is described, resulting in a percentage of correct scores.(3) See table 1 for an example of an quality indicator for pain intensity measured with the NPRS.

Table 1 Example of a Quality Indicator Monitoring the Process for Pain Intensity Measured With the Numeric Pain Rating Scale (NPRS)

Definition	The percentage of patients with low back pain who received physical therapist treatment and who completed the NPRS pretreatment and posttreatment to evaluate pain intensity.
Rationale	Pain reduction is an important goal after physical therapist treatment. Pain management is measured with the NPRS.
Numerator	The number of patients who received physical therapist treatment and who completed the NPRS pretreatment and posttreatment.
Denominator	All patients who received physical therapist treatment.
Specification	Pain intensity is measured in all patients using an 11-point NPRS, with 0 points being no pain at all and 10 points being unbearable pain.
Type of indicator	Process

For each of the four PROMs in the standard set, we defined four types of quality indicators:

- **Process indicator:** the percentage of patients with low back pain who had physical therapist treatment in which a pretreatment and posttreatment measurement was used. Example 1: in 60% of the patients, pain intensity was measured pretreatment and posttreatment with the NPRS.

- **Mean end scores:** the mean end score (with 95% CI) of patients with low back pain after physical therapy treatment. Example 2: the mean (SD) end score on physical function of patients measured with the QBPDS is 15 points.

- **Mean change scores:** the mean (SD) change score (with 95% CI) of patients with low back pain between pretreatment and posttreatment. Example 3: the mean change score in physical activity of patients measured with the PSFS is 3.2 points of improvement.

- **Minimal clinically important difference:** the percentage (with 95% CI) of patients who experienced a MCID between pretreatment and posttreatment to measure whether improvements in the outcome were clinically relevant. Example 4: in 70% (SD = 7%) of the patients reported an MCID on pain intensity measured with NPRS after treatment.

Monitoring the change score was not applicable for GPE-DV and was measured only posttreatment. Hence, in total we defined fifteen potential quality indicators. See Supplementary Appendix B for an extensive description of each defined potential quality indicator.

Describing the comparability and discriminability of the quality indicators

Sample size

To allow a valid comparison of indicator scores between physical therapists and practices, a rule of the thumb in multilevel analysis for general calculation is the 30/30 rule, i.e. 30 practices or physical therapists should include a minimum of 30 patients. (14, 15) We used this rule of thumb as threshold for estimating case-mix adjusted and stratified scores for each quality indicator. Descriptive statistics were used to determine whether thresholds for completion of PROMs were met for estimating (stratified) indicator scores. Physical therapists or physical therapy practices were excluded from the analysis for a specific quality indicator if they included less than 30 patients with NSLBP.

Comparability

We used linear and logistic multilevel analysis to compare outcomes on an aggregated level, i.e. the level of physical therapists and practices. In this study, patients are clustered through their physical therapists or clustered within practices.(16) To enhance the comparability of physical therapists or practices, the quality indicators were adjusted and stratified for explanatory variations in patient characteristics that influence the outcome, but are not under the control of the physical therapist or practice: (10)

a) Age, sex, chronicity and the baseline score of each PROM were used for adjustment in the multilevel analyses. The analysis aimed to explain the random intercept variance starting with estimating only an intercept and the random variation around the intercept (intercept-only model). Then, explanatory case-mix variables were added to the model (adjusted model), and the influence of the explanatory variables was evaluated by the amount of the random intercept variance that is explained. In this modelling procedure, the outcome of interest is not the regression coefficient of the explanatory variables but the amount of random intercept variance that is explained.(16)

b) The SBT was used to stratify patients for each quality indicator in low, medium and high-risk profiles.

For each physical therapist or practice in the case-mix adjusted multilevel analysis the mean scores were estimated with a 95% confidence interval (CI). For the analysis, we used PROMs measured at the beginning and end of the treatment.

Discriminative Ability

The Intraclass Correlation Coefficient (ICC) was calculated for each quality indicator to estimate the variation in outcomes between physical therapists or between practices. In this study, the ICC for physical therapists is defined as the variance

between physical therapists, divided by the total variance. The total variance is the summation of the variance between physical therapists and the variance within the physical therapists.(16) The ICC for physical therapy practices was defined accordingly by dividing the variance between practices by the summation of the variance between and the variance within physical therapy practices. In multilevel analysis, most ICCs are between .05 and .20 and ICCs >.10 can be interpreted as adequate, indicating that the quality indicator is able to discriminate outcomes between physical therapist or practices.(16, 17)

The ICC was also used to compare the intercept-only model with the adjusted model with explanatory variables (case-mix). For each final analysis, we stratified outcomes using the SBT, hence resulting in three (low-, medium- or high risk) multilevel analysis for each potential quality indicator.

Visual Representation of Indicator Scores

To present for each defined quality indicator the collection of mean outcomes of each participant (physical therapist- or practice level) in one graph, we used caterpillars plots.(17, 18) Caterpillar plots are regarded as user friendly and very suitable to visually display quality indicators.(13, 19) We used relative norms by presenting three colours, blue (95% CI significantly lower than average) purple (no significant 95% CI difference from average) and green (95% CI significantly higher than average). We used the plots to present the outcomes of the cohort data to participants in phase 2 of the study.

Phase 2: Selecting a core set of quality indicators

Participating physical therapists were purposefully invited for semi-structured interviews in focus groups. We intended to conduct at least three focus group meetings within every session 6-10 physical therapists. In addition, we conducted a focus group meeting with Dutch senior physical therapy researchers that were members of the development team of the revised Dutch physical therapy guideline for low back pain.

In these focus group meetings, participants were asked to choose a core set of quality indicators, selected from the 15 defined potential quality indicators. They reflected on the perceived added value of using the presented indicators in daily practice for quality improvement. In addition, the senior researchers were asked to comment on the set of quality indicators from a scientific perspective.

In all focus group meetings, we started with a visual representation of each potential quality indicator, both at the level of the physical therapist and of the practice in comparison with benchmark data. During the focus group meetings, participants compared, interpreted, and discussed the outcomes with their peers. The usability

was evaluated by the participants by interpreting the presented comparability and discriminability of all potential indicators. We also asked if the indicators were capable of making valid comparisons between physical therapists and practices. Furthermore, we were interested if the participants accepted the quality indicators as having added value for quality improvement initiatives. Finally, we asked them to select their preferred quality indicators for the core set.

After each focus group meeting, the chairman (AV) summarized the preferred quality indicators per outcome domain and asked if the group agreed with final proposed selection. The focus group meeting was ended when the majority of the participants reached consensus about the core set.

RESULTS

Phase 1: Defining and describing the comparability and discriminability of the quality indicators

Descriptive statistics

In table 2, descriptive statistics of included patients are presented for each SBT profile separately, including the number of physical therapists and practices. In total, 65,815 completed treatment episodes were provided by 1,900 physical therapists and 219 physical therapy practices. The patient characteristics per SBT profile in table 2 show that mean age, chronicity, treatment frequency and treatment weeks are similar across the profiles, except for the mean treatment frequency of SBT profile I in comparison with profiles II and III. Table 3 presents the provided data for each PROM at physical therapist- and practice level for the total population and stratified for each SBT profile. The change score of the NPRS and PSFS per SBT profile were similar, while the change score on the QBPDS fluctuated between 23.2 (SBT I) and 32.7 (SBT III) points. For the NPRS 22,740, the PSFS 17,540, the QBPDS 12,620 and the GPE 8,171 treatment episodes with repeated measurements were provided, see Supplementary Figure for an overview. The total number of included episodes (N) differed between the selected PROMs, since not all PROMs were registered in each patient record.

Table 2 Descriptive Statistics of Included Patients and Number of Participating Physical Therapists and Practices, Stratified per Start Back Tool Profile and for the Total Population

Number of treatment episodes in the dataset	**Total N = 72,226**	**SBT I N = 10,807**	**SBT II N = 5,310**	**SBT III N = 1,412**
Number of completed treatment episodes	Total N = 65,815	SBT I N = 9,580	SBT II N = 4,703	SBT III N = 1,300
Percentage of female patients	52.5%	47.3%	53.9%	55.5%
Mean age of patients (SD)	51.3 (16.9)	50.1 (16.3)	52.9 (16.1)	54.1 (16.2)
Percentage of patients with complaints <3 months	78.7%[a]	83.7%	88.6%	72.8%
Mean treatment frequency	7.0 (7.3)	5,6 (5,0)	7,1 (6,5)	7.6 (7,3)
Mean treatment weeks	9.2 (12.8)	7.2 (9.8)	8.3 (10.2)	8.7 (10.7)
Number of physical therapists who provided the data	1,009	687	579	313
Number of physical therapy practices that provided the data	219	179	181	130

SBT = Start Back Tool [a]Based on 34,460 patients in the dataset due to missing data. The total number of provided patient records, physical therapists and practices differs between the SBTs and total data, as an SBT was not registered in the data in all provided patient records.

Included cases in multilevel analysis

In total, 140 physical therapists (out of 1,009) and 85 physical therapy practices (out of 219) fulfilled the 30/30 rule for at least one PROM with successful follow-up measurement (i.e. at least two scores on the PROM in patients that had at least two visits) and could be included in the multilevel analysis. For modeling outcomes of pain intensity using the NPRS, a total of 13,096 (30% of total) treatment episodes were provided. For modeling physical activity measured with the PSFS, 106 physical therapists and 83 physical therapy practices provided a total of 10,363 (25% of total) treatment episodes. For modeling physical functioning measured with the QBPDS, 88 physical therapists and 55 physical therapy practices provided a total of 9,437 (15% of total) treatment episodes. For modeling perceived effect measured with the GPE-DV 53 physical therapists and 48 physical therapy practices, provided a total of 7,121 (11% of total) treatment episodes. The 30/30 inclusion rule was not reached for each PROM and associated outcome domain per SBT subgroup

Table 3 Descriptive Statistics of Mean PROM Scores (Unadjusted) for the Total Population and Stratified per SBT Profilea

	NPRS	PSFS	QBPDS	GPE-DV[b]
Percentage of female patients	51.9%	52.1%	51.1%	50.2%
Mean age patients (SD)	51.9 (16.8)	50.8 (16.7)	51.5 (16.4)	51.4 (16.4)
Percentage of patients with complaints <3 months	79.2%	78.9%	79.3%	80.8%
Number treatment episodes with baseline measures	44,251	46,016	26,142	-
Mean baseline score (SD)	6.3 (1.8)	6.8 (1.92)	36.8 (19.1)	-
Included patients with end scores	22,740	17,540	12,620	8,171
Mean end score	3.3 (2.8)	2.8 (2.4)	12.6 (15.3)	
Mean change Tend-T0 (SD)	30 (3.1)	4.2 (2.6)	26.1 (19.2)	1.8[b]
MCID of each PROM (reference)	>2 points *(46, 47)*	>2 points *(48)*	>20 points *(46, 49)*	Score 1 or 2 *(50)*
Percentage patients improved[c]	71.5%	83.8%	57.5%	84.8%
Percentage patients stabilized[c]	20%	14.9%	41.7%	15.1%
Percentage patients deteriorated[c]	8.5%	1.3%	0.8%	0.1%
Number of physical therapists that provided data	935	967	879	543
Number of practices that provided data	202	210	202	145
Number of treatment episodes profile SBT I	7,177	7,485	6,699	2,448
SBT I: mean change Tend-T0 (SD)	4.0 (2.2)	4.4 (2.3)	23.2 (17.4)	-
Number of treatment episodes profile SBT II	4,189	4,215	3,799	1,718
SBT II: mean change Tend-T0 (SD)	4.3 (2.3)	4.7 (2.6)	28.6 (19.3)	-
Number of treatment episodes profile SBT III	1,195	1,196	1,044	294
SBT III: mean change Tend-T0 (SD)	4.1 (2.5)	4.5 (3.7)	32.7 (23.4)	-

[a]GPE-DV = Global Perceived Effect - Dutch Version; MCID = minimal clinically important difference; NPRS = Numeric Pain Rating Scale; PROM = patient-reported outcome measures; PSFS = Patient Specific Functional Scale; QBPDS = Quebec Back Pain Disability Scale; SBT = Start Back Tool [b]GPE is monitored only at the end of the treatment.
[c]Improved, stabilized, or deteriorated based on MCID cutoff points.

Comparability & discriminative ability

Table 4 shows that the ICCs fluctuated between 0.08 and 0.30 and was therefore judged as adequate to discriminate between physical therapists or practices.(17) The adjusted model was used for the visual representation of the quality indicators in the focus group interviews.

Table 4 Intraclass Correlation Coefficients for Intercept-Only Model and Adjusted Model for Change, End, and MCID Scores for Each PROM[a]

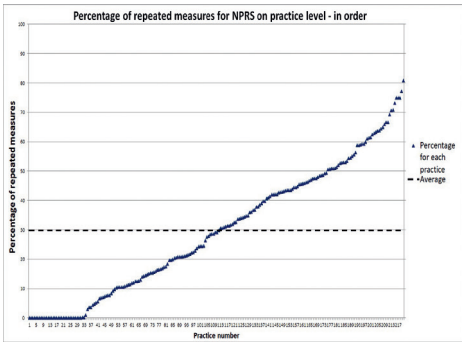
	Physical therapist level		Practice level	
	Intercept-only model	**Adjusted model**	**Intercept-only model**	**Adjusted model**
NPRS end score	0.12	0.12	0.08	0.08
NPRS change score	0.08	0.12[b]	0.06	0.08[b]
NPRS MCID	0.11	0.12[b]	0.09	0.10[b]
PSK end score	0.12	0.12	0.09	0.09
PSK change score	0.13	0.16[b]	0.08	0.11[b]
PSK MCID	0.16	0.17[b]	0.11	0.12[b]
QBPDS end score	0.15	0.16[b]	0.15	0.15
QBPDS change score	0.20	0.30[b]	0.12	0.21[b]
QBPDS MCID[c]	0.10	0.13[b]	0.10	0.13[b]
GPE-DV end score	0.18	0.18	0.10	0.10
GPE-DV MCID	0.14	0.14	0.13	0.13

[a] With repeated measures for the total population on practice level and physical therapist level that provided >30 patients. GPE-DV = Global Perceived Effect – Dutch Version; MCID = minimal clinically important difference; NPRS = Numeric Pain Rating Scale; PSFS: Patient Specific Functional Scale; QBPDS = Quebec Back Pain Disability Scale; SBT = Start Back Tool. [b] Increasement of the ICC in comparison with the intercept-only model with adjustment for case-mix variables age, sex, chronicity, and begin score of the PROM. [c] For the multilevel analysis of the QBPDS, an MCID of >20 points was used.

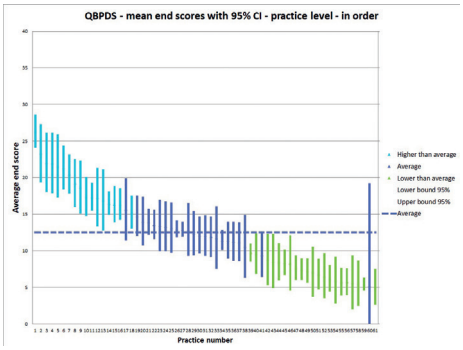
Visual representation of indicator scores

Examples of caterpillar plots from the quality indicators listed above are presented in the Figure. In line with the ICC (range = 0.08 - 0.30), the Supplementary Figure and the Figure show adequate variation of treatment outcomes among physical therapist practices. Roughly one third of the physical therapist practices were divided in significantly lower than average, no-difference, or higher than average based on 95% CI. See Supplementary Appendix C for a visual representation of each defined potential quality indicator.

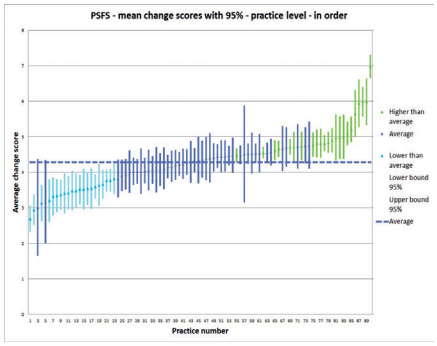
Process indicator[a]:



End score outcome indicator[b]:



Change score outcome indicator[c]:



Experienced MCID score outcome indicator[d]:

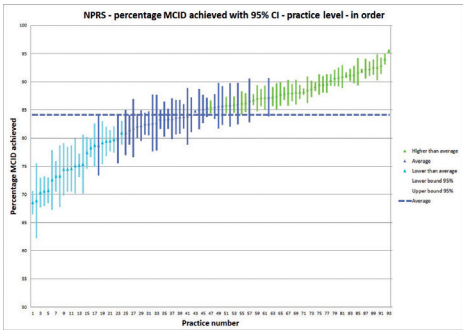


Figure [a] The percentage of patients with low back pain who received physical therapist treatment in which a pre and/or post NPRS was used [b] The mean end score with 95% CI on QBPDS of patients with nonspecific low back pain after physical therapist treatment [c] The mean change score with 95% CI on PSFS of patients with nonspecific low back pain between pretreatment and posttreatment [d] The percentage with 95% CI of patients with non-specific low back pain who experienced a MCID on the NPRS between pretreatment and posttreatment. To measure whether improvements in the outcome were clinically relevant

Phase 2: Selecting a core set of quality indicators

Nineteen physical therapists and four senior researchers participated in four focus group interviews. The mean age was 37 years (range = 25 – 57), the average work experience was 11 years (range = 1 – 31). In general, the participants accepted the quality indicators for having added value in daily practice as quality improvement tools. The relative norms as presented (see Figure) were interpreted as user friendly and easy to read. Participants mentioned that using the quality indicators can be a good opportunity to stimulate transparency of variation in outcomes between practices, but emphasized that for learning goals it is very important to create a safe environment for physical therapists when comparing outcomes. Furthermore, physical therapists expressed that, when comparing outcomes with their peers, they were interested in additional information, e.g. treatment frequency and recurrences of NSLBP.

The majority of each focus group preferred the process indicators for routine measurement of the NPRS and PSFS, the pre-post treatment change scores of the NPRS, PSFS, QBPDS, and the MCID based on the post-treatment score of the GPE-DV for inclusion in the core set. Hence, the potential indicators with end scores and MCID were not selected for the NPRS, PSFS and QBPDS. The final core set of quality indicators is presented in table 5.

Table 5 Selected Core Set of Quality Indicators That Is Accepted by Stakeholders on Usability and Perceived Added Value as Quality Improvement Tools

Type of indicator	PROM	Domain	Quality indicator description	Mean/ percentage (SD)[b]	Range[b]
Process	NPRS	Pain intensity	The percentage of patients with nonspecific low back pain who received physical therapy treatment and who completed the NPRS pretreatment and posttreatment to evaluate pain intensity	30%	0-80%
Process	PSFS	Physical activity	The percentage of patients with nonspecific low back pain who received physical therapy treatment and who completed the PSFS pretreatment and posttreatment with the PSFS to evaluate physical activity	25%	0-75%
Change	NPRS	Pain intensity	The mean change score + 95% CI of patients with nonspecific low back pain who received physical therapy treatment and are pre- and post-measured with the NPRS to evaluate pain intensity	3.8 (0.6)	2.3 – 5.8
Change	PSFS	Physical activity	The mean change + 95% CI score of patients with nonspecific low back pain who received physical therapy treatment and are pre- and post-measured with the PSFS to evaluate physical activity	4.2 (0.7)	2.7 – 7.0
Change	QBPDS	Physical function	The mean change score + 95% CI of patients with nonspecific low back pain who received physical therapy treatment and are pre- and post-measured with the QBPDS to evaluate physical functioning	24.6 (6.6)	11.2 – 43.4
MCID	GPE-DV	Perceived effect	The percentage of patients with nonspecific low back pain who received physical therapy treatment and experienced a MCID effect on the GPE-DV	88%	58-96%

[a]GPE-DV = Global Perceived Effect—Dutch Version; MCID = minimal clinically important difference; NPRS = Numeric Pain Rating Scale; PROM = patient reported outcome measure; PSFS = Patient Specific Functional Scale; QBPDS = Quebec Back Pain Disability Scale. [b] The mean/percentage and range are the outcomes of the cohort data that were used for describing and selecting the quality indicators for the core set.

DISCUSSION

Out of 15 defined potential outcome-based quality indicators, stakeholders selected a final core set of six quality indicators. The core set contains two process indicators, the percentage of completed NPRS for measuring pain intensity and the percentage of completed PSFS for measuring physical activity; and four outcome indicators, including change scores for the NPRS, PSFS, QBPDS for measuring pain intensity, physical activity and physical functioning, respectively, and the GPE for measuring perceived effect using the MCID. All previously selected PROMs and associated outcome domains were included in the core set.(7) The discriminability of the outcome-based quality indicators between physical therapists and physical therapy practices was adequate based on the ICCs. The outcome-based quality indicators were accepted by stakeholders for having added value in daily practice and for quality improvement purposes. To our knowledge, this is the first core set of outcome-based quality indicators developed for primary care physical therapists and practices treating patients with NSLBP.

The findings in the present study are supported with the work of other studies in this area. For example the study of Hildon et al. (2015), that described views of patients and clinicians on comparing quality indicators for providers of surgery, underlined the value of comparison of outcomes between providers for quality improvement. (20) The visual representation of caterpillars plots and the selection of the change scores and MCID for defining quality indicators based on PROMs has been used in other studies.(5, 13, 19)

The main difference with the present study is that our study explicitly focused on collecting and interpreting outcomes of care, while other studies in mainly focused on process indicators, for example by monitoring guideline adherence, the quality of the clinical reasoning process and shared decision making.(21-26)

Also, although not all physical therapists routinely collected enough data, the amount of participating physical therapists and provided cohort data is significantly higher than in other studies.(25, 26). With physical therapists who did routinely collected data and participated in focus group meetings for selection of the core set, engagement of end-users was optimally stimulated. We therefore believe that the usability and added value of the core set for daily practice is guaranteed, but that more effort is needed to implement the quality indicators in clinical practice.

The selected outcome domains and associated PROMs in the core set, i.e. pain intensity, physical activity, physical functioning and perceived effect, are partly in line with other research regarding the development of core outcome sets (COS) for patients

with NSLBP.(27-31) However, these COS are developed for clinical trial purposes and have not been tested for their relevance and feasibility in daily practice.(7)

To increase the usability of the indicators, we chose to include all available case-mix variables for adjustment in the multilevel analysis, in order to enhance the acceptance of the quality indicators as a quality improvement tool. Excluding these case-mix variables might lead physical therapists to attribute differences in outcomes to these factors (e.g. age of the patient or chronicity of the LBP) instead of to their own treatment decisions.

Implications for research and practice

As presented, physical therapists and physical therapist practices, who sample routinely enough data, can use this core set for monitoring and comparing treatment outcomes in research and daily practice in order to evaluate and stimulate quality of care. To develop a full picture of the validity and usefulness of the core set of outcome-based quality indicators, further research will be needed in which quality indicators are measured over time. Quality indicator scores can then be used for longitudinal evaluation and monitoring of achievements of physical therapists and practices using outcomes of patients.

An example of an approach for daily practice is comparing outcomes and receiving feedback from supervisors or colleagues in peer-review assessments.(32, 33) In such an approach, it is important to formulate explicit targets and action plans to enhance quality improvement initiatives.(32) A plan-do-study-act (PDSA) cycle is a way to provide a structure for iterative testing and evaluation of change to improve quality of care.(34) When using the PDSA cycle, physical therapists and physical therapy practices can define action plans that are based on their own specific learning goals. The quality indicators can play a key role in this process, thus leading to a substantial improvement in the care for patients with NSLBP.

As also mentioned by participating physical therapists in our focus groups, an important condition for comparing outcomes between peers is creating a safe environment to share and discuss real-world data.(35). A safe environment for learning purposes stimulates the possibility to learn from each other, to try different treatment approaches and to learn from mistakes, and thus to increase outcomes for future patients based on the interpretation of outcomes. Recently, Dutch physical therapy associations in physical therapy have developed manuals that guide peer assessment meetings in comparing real-world outcomes data as an instrument for continuous quality improvement.(36, 37) The next step is to investigate the impact of these meetings on quality of care.

Limitations

The most relevant limitation of our study is that many physical therapists did not reach the threshold of 30 treatment episodes with repeated measurements of PROMs, and they were therefore excluded for the multilevel analysis, and therefore from the comparison between physical therapists and practices. Moreover, the missing values on the SBT prohibited us to perform the multilevel analysis with the subgroups based on the SBT, while these subgroups were judged as clinically relevant by the physical therapists. This was confirmed by the descriptive statistics (see table 3) showing differences in baseline and change scores between the subgroups.

There are several possible reasons that many physical therapists did not provide >30 patients with NSLBP, including low motivation for data sampling or having problems with data sampling routines in physical therapists, low motivation or missing skills to answer the questionnaires in patients, and technical issues with data extraction from the EHR to the national registries.(2) Physical therapists in the Netherlands register their data via various EHR vendors that are responsible for providing data to national data registries.(2) Each EHR has its own data infrastructure, interface and privacy policy, which makes it complex to standardize outcomes in national data registries. Hence, standardizing outcomes of patients for each EHR vendor is time-consuming and requires collaboration of all stakeholders, including researchers, physical therapists, policy makers and most importantly the EHR vendors itself to solve technical issues.(2) For example, technical issues can be identified and addressed by conducting end-to-end validations to investigate whether outcomes provided by physical therapists and patients are correctly documented in a national registry.

To implement the use of patient-reported outcome-based quality indicators in daily practice, an analysis of implementation determinants hindering or facilitating the data sampling process is necessary. These determinants are needed to choose strategies to improve the process from data sampling to data evaluation and to change practice routines.(38) It is obvious that not a single implementation strategy will be sufficient while differences between practices are large, so the implementation strategies should be tailored to determinants relevant in individual practices. However, studies in the past showed that using implementation strategies considerably increases the registration in physical therapist practice. (2, 39, 40) Moreover, a qualitative study in physical therapy patients showed that they are motivated to use PROMs if physical therapists explain the added value.(41) Therefore, we conclude that the next step should be focused on tailored implementation strategies to enable, motivate and train patients and physical therapists to structurally register PROMs by solving technical and organizational issues with data sampling by patients and data extraction from EHRs.

Another limitation is that our study did not use a formal voting system to reach consensus in the selection of the core set, as is common in Delphi studies for

developing quality indicators.(42) However, the majority in each of the four focus groups selected the same outcome-based quality indicators. We think the absence of a formal voting system did not affect the selection of the core set of indicators.

In the current study we used convenience sampling for recruiting physical therapists. This may potentially have led to selection bias of Dutch physical therapists enthusiastic for participation in a study collecting and comparing their outcomes with their peers.

Indicators using the MCID were expressed as the percentage (with 95% CI) of patients who experienced a MCID between pretreatment and posttreatment to measure whether improvements in the outcome were clinically relevant. Natural history and regression to the mean are potential confounders that may influence indicator scores. However, we did not aim to perform an effectiveness study in a controlled setting, and we *explored* for the explained variation of real-world outcomes on the level of physical therapist practice. Still, when using MCID scores in quality indicators, the minimal detectable change (MDC) needs also be taken into account because the MCID scores may be smaller than the minimal detectable changes (MDCs) of investigated PROMs. In the current study, the MDC of included measures where all within or the same as the MCID, for the PSFS: 1.4 (43) (MCID 2 points) QBPDS 15.8 points (44) (MCID 20 points) NPRS 2 points (45) (MCID 2 points).

Future work is required to identify more patient characteristics that could influence treatment outcomes of physical therapists or physical therapy practices. For example, socioeconomic status (SES) or recurrences of episodes in the same patient could be used as a case-mix variable in the multilevel analysis

Conclusion

This study has defined, described and selected a core set of outcome-based quality indicators based on cohort data and consensus. The core set was accepted by users and stakeholders for having added value for daily practice in physical therapy primary care and was found useful for quality improvement initiatives. Further studies should focus on further tailored implementation strategies that stimulate the use of the core set of outcome-based quality indicators, to register PROMs routinely for monitoring the quality of physical therapy care, and to use the indicators in plan-do-study-act cycles by evaluating specific improvement goals at the level of physical therapists or physical therapy practices.

Acknowledgments

The authors acknowledge the following individuals: Juliette Cruijsberg and Janine Liefers, research members; Francois Maissan, Bob van den Meiracker, Jasper Bier, and Harald Miedema, senior researchers; and all physical therapists who participated and provided data or participated in the focus group meetings.

REFERENCES

1. Westby MD, Klemm A, Li LC, Jones CA. Emerging role of quality indicators in physical therapist practice and health service Delivery. Physical therapy. 2016;96(1):90-100.
2. Meerhoff GA, van Dulmen SA, Maas MJM, Heijblom K, Nijhuis-van der Sanden MWG, Van der Wees PJ. Development and Evaluation of an Implementation Strategy for Collecting Data in a National Registry and the Use of Patient-Reported Outcome Measures in Physical Therapist Practices: Quality Improvement Study. Physical Therapy. 2017;97(8):837-51.
3. van der Wees PJ, Verkerk EW, Verbiest MEA, Zuidgeest M, Bakker C, Braspenning J, et al. Development of a framework with tools to support the selection and implementation of patient-reported outcome measures. Journal of Patient-Reported Outcomes. 2019;3(1):75.
4. Cella D, Hahn EA, Jensen SE, Butt Z, Nowinski CJ, Rothrock N, editors. Methodological issues in the selection, administration and use of patient-reported outcomes in performance measurement in health care settings. NQF (NQF), Washingtdon, DC: National Quality Forum (NQF); 2012.
5. Patient Reported Outcomes (PROs) in Performance Measurement: National Quality Form (NQF); 2012.
6. Campbell SM, Braspenning J, Hutchinson A, Marshall MN. Research methods used in developing and applying quality indicators in primary care. BMJ. 2003;326(7393):816-9.
7. Verburg AC, van Dulmen SA, Kiers H, Nijhuis-van der Sanden MWG, van der Wees PJ. Development of a standard set of outcome measures for non-specific low back pain in Dutch primary care physiotherapy practices: a Delphi study. European Spine Journal. 2019;28(7):1550-64.
8. Weldring T, Smith SM. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). Health Serv Insights. 2013;6:61-8.
9. Kessel Pv, Triemstra, M., Boer, D. de. Handreiking voor het meten van kwaliteit van zorg met Patient Reported Outcome Measures 2014 [Available from: https://nivel.nl/sites/default/files/bestanden/Handreiking-meten-kwaliteit-zorg-met-Patient-Reported-Outcome-Measures.pdf.
10. Greenhalgh J, Dalkin S, Gooding K, Gibbons E, Wright J, Meads D, et al. Health Services and Delivery Research. Functionality and feedback: a realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care. Southampton (UK): NIHR Journals Library; 2017.
11. Hill JC, Dunn KM, Lewis M, Mullis R, Main CJ, Foster NE, et al. A primary care back pain screening tool: identifying patient subgroups for initial treatment. Arthritis and rheumatism. 2008;59(5):632-41.
12. Staal J, Hendriks E, Heijmans M, Kiers H, Lutgers-Boomsma G, Rutten G. KNGF-richtlijn Lage rugpijn Verantwoording en toelichting. KNGF. 2013;21.
13. Hildon Z, Neuburger J, Allwood D, van der Meulen J, Black N. Clinicians' and patients' views of metrics of change derived from patient reported outcome measures (PROMs) for comparing providers' performance of surgery. BMC health services research. 2012;12:171.
14. Ali A, Ali S, Khan SA, Khan DM, Abbas K, Khalil A, et al. Sample size issues in multilevel logistic regression models. PloS one. 2019;14(11):e0225427-e.

15. Koolman X, Zuidgeest M, Visser J, Appelman M. Indicatorstandaard 2.0: Methodologische criteria voor de ontwikkeling van betrouwbare kwaliteitsindicatoren in de zorg.2012.

16. Twisk JWR. Applied Mixed Model Analysis: A Practical Guide. 2 ed. Cambridge: Cambridge University Press; 2019.

17. Boer D, Triemstra M, Koopman L, Zuidgeest M. Tool Analyses ; voor het vergelijken van patiëntervaringen tussen zorgaanbieders | versie 1.0 2016 [Available from: https://www.zorginzicht.nl/binaries/content/documents/zorginzicht/ontwikkeltools/prom-toolbox/prom-cyclus-stap-5-definieren-indicator/prom-cyclus-stap-5-definieren-indicator/files/3/Tool-Analyse-voor-het-vergelijken-van-pati%C3%ABntervaringen-tussen-zorgaanbieders.pdf.

18. Goldstein H, Michael JRH. The Graphical Presentation of a Collection of Means. Journal of the Royal Statistical Society Series A (Statistics in Society). 1995;158(1):175-7.

19. Hildon Z, Allwood D, Black N. Making data more meaningful: patients' views of the format and content of quality indicators comparing health care providers. Patient education and counseling. 2012;88(2):298-304.

20. Hildon Z, Allwood D, Black N. Patients' and clinicians' views of comparing the performance of providers of surgery: a qualitative study. Health Expect. 2015;18(3):366-78.

21. Petersson IF, Strombeck B, Andersen L, Cimmino M, Greiff R, Loza E, et al. Development of healthcare quality indicators for rheumatoid arthritis in Europe: the eumusc.net project. Ann Rheum Dis. 2014;73(5):906-8.

22. Westby MD, Marshall DA, Jones CA. Development of quality indicators for hip and knee arthroplasty rehabilitation. Osteoarthritis and cartilage. 2018;26(3):370-82.

23. Peter WF, Hurkmans EJ, Wees PJvd, Hendriks EJM, Bodegom-Vos Lv, Vlieland TPMV. Healthcare Quality Indicators for Physiotherapy Management in Hip and Knee Osteoarthritis and Rheumatoid Arthritis: A Delphi Study. Musculoskeletal Care. 2016;14(4):219-32.

24. Scholte M, Neeleman-van der Steen CW, Hendriks EJ, Nijhuis-van der Sanden MW, Braspenning J. Evaluating quality indicators for physical therapy in primary care. International journal for quality in health care : journal of the International Society for Quality in Health Care. 2014;26(3):261-70.

25. Oostendorp RA, Rutten GM, Dommerholt J, Nijhuis-van der Sanden MW, Harting J. Guideline-based development and practice test of quality indicators for physiotherapy care in patients with neck pain. J Eval Clin Pract. 2013;19(6):1044-53.

26. Oostendorp RAB, Elvers JWH, van Trijffel E, Rutten GM, Scholten-Peeters GGM, Heijmans M, et al. Relationships Between Context, Process, and Outcome Indicators to Assess Quality of Physiotherapy Care in Patients with Whiplash-Associated Disorders: Applying Donabedian's Model of Care. Patient Prefer Adherence. 2020;14:425-42.

27. Chiarotto A, Boers M, Deyo RA, Buchbinder R, Corbin TP, Costa LOP, et al. Core outcome measurement instruments for clinical trials in nonspecific low back pain. Pain. 2018;159(3):481-95.

28. Chiarotto A, Deyo RA, Terwee CB, Boers M, Buchbinder R, Corbin TP, et al. Core outcome domains for clinical trials in non-specific low back pain. European Spine Journal. 2015;24(6):1127-42.

29. Clement RC, Welander A, Stowell C, Cha TD, Chen JL, Davies M, et al. A proposed set of metrics for standardized outcome reporting in the management of low back pain. Acta orthopaedica. 2015;86(5):523-33.

30. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, et al. Outcome measures for low back pain research. A proposal for standardized use. Spine. 1998;23(18):2003-13.

31. Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. Spine. 2000;25(24):3100-3.

32. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. The Cochrane database of systematic reviews. 2012(6):Cd000259.

33. Maas MJM, Driehuis F, Meerhoff GA, Heerkens YF, van der Vleuten CPM, Nijhuis-van der Sanden MWG, et al. Impact of Self- and Peer Assessment on the Clinical Performance of Physiotherapists in Primary Care: A Cohort Study. Physiother Can. 2018;70(4):393-401.

34. Taylor MJ, McNicholas C, Nicolay C, Darzi A, Bell D, Reed JE. Systematic review of the application of the plan-do-study-act method to improve quality in healthcare. BMJ Qual Saf. 2014;23(4):290-8.

35. Maas MJ, Nijhuis-van der Sanden MW, Driehuis F, Heerkens YF, van der Vleuten CP, van der Wees PJ. Feasibility of peer assessment and clinical audit to self-regulate the quality of physiotherapy services: a mixed methods study. BMJ open. 2017;7(2):e013726.

36. Maas MJ, de Leur M, Wees PJvd, Kiers H. Peer review - meetinformatie gebruiken om te leren en verbeteren 2020 [Available from: https://www.keurmerkfysiotherapie.nl/wp-content/uploads/2020/06/peer-review-2020-thema-1.-data-1.0.pdf.

37. Maas MJ, van Poelgeest E, Weymans C, Wees PJvd. Module 3 feedback op proces en uitkomsten KNGF. 2018.

38. van Achterberg T, Schoonhoven L, Grol R. Nursing implementation science: how evidence-based nursing requires evidence-based implementation. J Nurs Scholarsh. 2008;40(4):302-10.

39. Baker R, Camosso-Stefinovic J, Gillies C, Shaw EJ, Cheater F, Flottorp S, et al. Tailored interventions to address determinants of practice. The Cochrane database of systematic reviews. 2015;2015(4):Cd005470.

40. Bosch M, van der Weijden T, Wensing M, Grol R. Tailoring quality improvement interventions to identified barriers: a multiple case analysis. J Eval Clin Pract. 2007;13(2):161-8.

41. Meerhoff GA, van Dulmen SA, Maas MJM, Bakker-Jacobs A, Nijhuis-Van der Sanden MWG, van der Wees PJ. Exploring the perspective of patients with musculoskeletal health problems in primary care on the use of patient-reported outcome measures to stimulate quality improvement in physiotherapist practice; a qualitative study. Physiotherapy theory and practice. 2019:1-12.

42. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and Reporting the Delphi Method for Selecting Healthcare Quality Indicators: A Systematic Review. PLOS ONE. 2011;6(6):e20476.

43. Maughan EF, Lewis JS. Outcome measures in chronic low back pain. European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society. 2010;19(9):1484-94.

44. Demoulin C, Ostelo R, Knottnerus JA, Smeets RJ. Quebec Back Pain Disability Scale was responsive and showed reasonable interpretability after a multidisciplinary treatment. J Clin Epidemiol. 2010;63(11):1249-55.

45. Childs JD, Piva SR, Fritz JM. Responsiveness of the numeric pain rating scale in patients with low back pain. Spine. 2005;30(11):1331-4.

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

47. Farrar JT, Young JP, Jr., LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain. 2001;94(2):149-58.

48. van Dulmen SA, van der Wees PJ, Bart Staal J, Braspenning JCC, Nijhuis-van der Sanden MWG. Patient reported outcome measures (PROMs) for goalsetting and outcome measurement in primary care physiotherapy, an explorative field study. Physiotherapy. 2017;103(1):66-72.

49. Schoppink LE, van Tulder MW, Koes BW, Beurskens SA, de Bie RA. Reliability and validity of the Dutch adaptation of the Quebec Back Pain Disability Scale. Phys Ther. 1996;76(3):268-75.

50. Ostelo RW, de Vet HC. Clinically important outcomes in low back pain. Best practice & research Clinical rheumatology. 2005;19(4):593-607.

51. McCaffery M. Using the 0-to-10 pain rating scale. The American journal of nursing. 2001;101(10):81-2.

52. Froud R, Fawkes C, Foss J, Underwood M, Carnes D. Responsiveness, Reliability, and Minimally Important and Minimal Detectable Changes of 3 Electronic Patient-Reported Outcome Measures for Low Back Pain: Validation Study. J Med Internet Res. 2018;20(10):e272-e.

53. Beurskens AJ, de Vet HC, Koke AJ, Lindeman E, van der Heijden GJ, Regtop W, et al. A patient-specific approach for measuring functional status in low back pain. Journal of manipulative and physiological therapeutics. 1999;22(3):144-8.

54. Barten JA, Pisters MF, Huisman PA, Takken T, Veenhof C. Measurement properties of patient-specific instruments measuring physical function. J Clin Epidemiol. 2012;65(6):590-601.

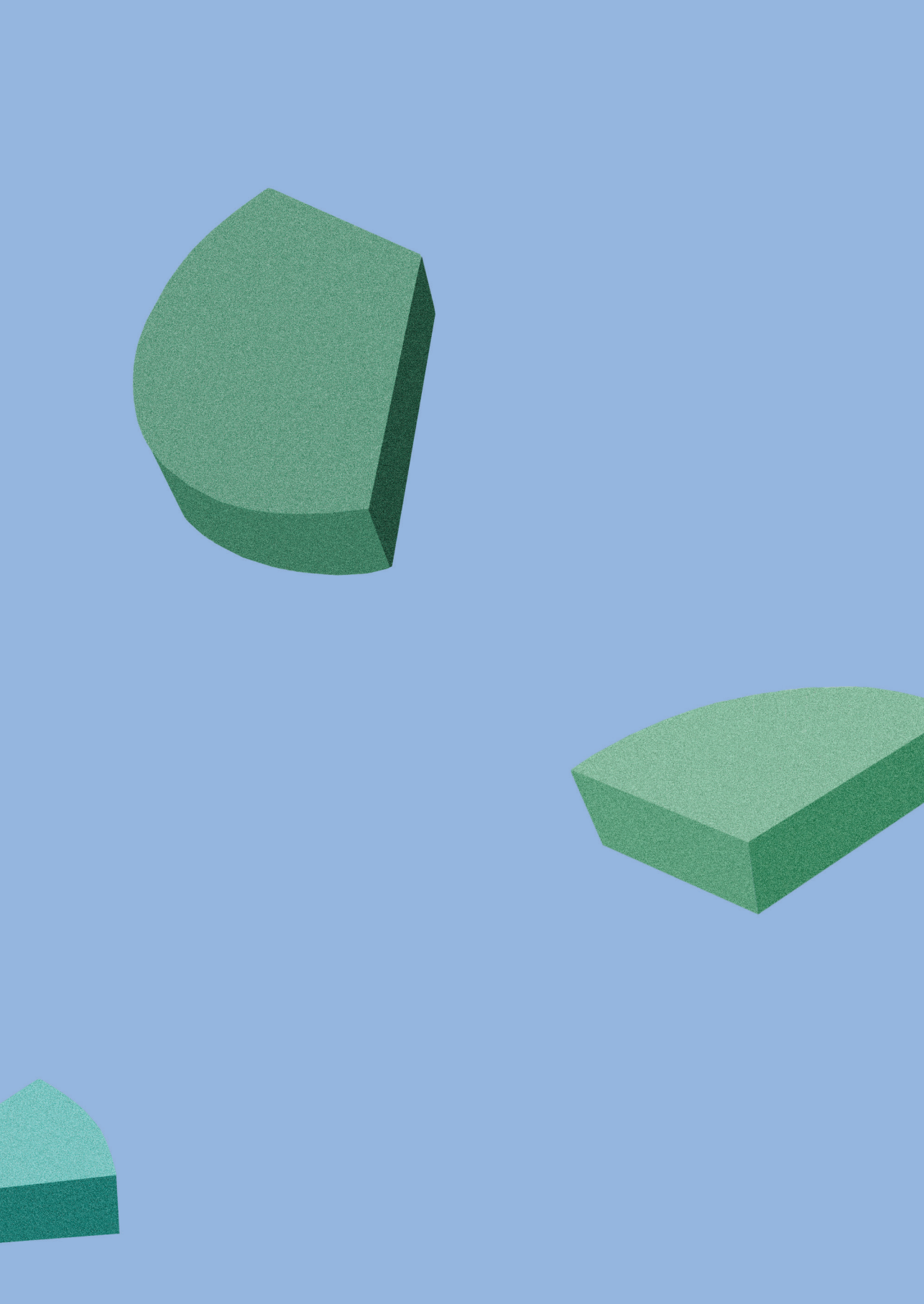
55. Hudak PL, Wright JG. The characteristics of patient satisfaction measures. Spine. 2000;25(24):3167-77.

56. Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. J Clin Epidemiol. 2010;63(7):760-6.e1.

57. Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, et al. The Quebec Back Pain Disability Scale. Measurement properties. Spine. 1995;20(3):341-52.

58. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. Physiotherapy. 1980;66(8):271-3.

59. Chiarotto A, Maxwell LJ, Terwee CB, Wells GA, Tugwell P, Ostelo RW. Roland-Morris Disability Questionnaire and Oswestry Disability Index: Which Has Better Measurement Properties for Measuring Physical Functioning in Nonspecific Low Back Pain? Systematic Review and Meta-Analysis. Phys Ther. 2016;96(10):1620-37.



CHAPTER 5

A practice test and selection of a core set of outcome-based quality indicators in Dutch primary care physical therapy for patients with COPD

A cohort study

Arie C. Verburg
Simone A. van Dulmen
Henri Kiers
Maria W.G. Nijhuis-van der Sanden
Philip J. van der Wees

Accepted for publication in European Respiratory Journal Open Research

ABSTRACT

Aim to estimate the comparability and discriminability of outcome-based quality indicators by performing a practice test in Dutch physical therapy primary care, and to select a core set of outcome-based quality indicators that are well-accepted by physical therapists based on their perceived added value as a quality improvement tool.

Methods First, a list of potential quality indicators was defined, followed by the determination of the comparability (case-mix adjusted multi-level analysis) and discriminability (intraclass correlation coefficient (ICC)). Second, focus group meetings were conducted with stakeholders (physical therapists and senior researchers) to select a core set of quality indicators.

Results Overall, 229 physical therapists from 137 practices provided 2651 treatment episodes. Comparability: in 10 of the 11 case-mix adjusted models, the ICC increased compared with the intercept-only model. Discriminability: the ICC ranged between 0.01 and 0.34, with five of the 11 ICCs being > 0.10. The majority of physical therapists in each focus group preferred the inclusion of seven quality indicators in the core set, including three process and four outcome indicators based upon the Six-Minute-Walk-Test (6MWT), the Clinical COPD Questionnaire (CCQ), and the determination of quadriceps strength using a hand-held dynamometer (HHD).

Conclusion This is the first study that describes the comparability and discriminability of the outcome-based quality indicators selected for patients with COPD treated in primary care physical therapy practices. Future research should focus on increasing data collection in daily practice and on the development of tangible methods to use as the core set of a quality improvement tool.

BACKGROUND

The routine use of outcome measures can play an important role in improving healthcare quality;(1) for example, they can enable the comparison of providers' performances to stimulate improvement initiatives.(2) A fundamental prerequisite of the use of outcome measures is the collection, aggregation, and comprehensive understandable presentation of data.(1) Using quality indicators may stimulate the routine data collection of patient-reported outcome measures (PROMs) by healthcare providers. Quality indicators can be used on an aggregated level to show changes in clinical practice over time.(1, 3, 4)

In a previous study, we developed a standard set of outcome domains and associated measures, including PROMs and physical performance measures, for patients with chronic obstructive pulmonary diseases (COPD) in primary care physical therapy practice.(4) It is still unclear which quality indicators can be selected from the standard set and which quality indicators have perceived added value as quality improvement tools for such patients, however.

In this study we focused on outcome-based quality indicators chosen from the standard set of PROMs and physical performance measures for patients with COPD.(5) PROMs are often combined with other clinician-assessed, impairment-based or physical performance–based measures, such as the Six-Minute Walk Test (6MWT), to provide a more complete interpretation of patient outcomes.(6)

Currently, most strategies for the development of quality indicators are based on an evidence-based consensus between stakeholders in procedures, such as the RAND/UCLA Delphi procedure.(7-9) This is true for recommendations for clinical practice guidelines too, such as the recently published Dutch clinical practice guideline (CPG) for primary care physical therapists treating patients with COPD,(10) and can provide an important basis for the development of quality indicators.(11) In addition, a practice test, including the collection of real-world data prior to selection, is an essential step for the evaluation of the comparability, discriminability, and feasibility of potential quality indicators in daily practice.(9, 12) A practice test can support the usefulness and feasibility of quality indicators in daily practice and gain insight into the psychometric properties of outcome-based quality indicators. (3, 11, 13) Although to our knowledge, no specific definition of a practice test is reported in previous research, there are several examples of using a practice test in development of quality indicators. (11, 13, 14) Such an example is the study of Meerhoff et al. 2021 in which a practice test was conducted to explore the reliability, validity and discriminability of patient reported outcomes for the development of quality indicators in patients with non-specific low back pain.(13) We defined comparability as the extent to which the

quality indicator is comparable between practices, and discriminability as the extent to which the quality indicator is able to discriminate between practices.

Here, we develop outcome-based quality indicators for patients with COPD in physical therapy primary care. The aims of this study are therefore:

a) To estimate the comparability and discriminability of outcome-based quality indicators included in a previously selected standard set of measures;

b) To select a core set of outcome-based quality indicators that is well-accepted by physical therapists based on the perceived added value of this core set as a quality improvement tool.

METHODS

Design

In this mixed methods study, we used a sequential explanatory design taking a previously selected standard set of outcome domains and measures as the basis for defining and selecting a core set of quality indicators. The standard set was developed in two consecutive steps between February 2018 and April 2020,(4) and was registered on the Core Outcome Measures in Effectiveness Trials (COMET) website.(15) In phase 1 of the present study, potential quality indicators were defined, and we estimated their comparability and discriminability with prospectively collected cohort data between February 2018 and December 2019. To enhance the comparability, we adjusted for differences in patient characteristics using a case-mix correction. Furthermore, we calculated whether the quality indicator was able to discriminate the outcomes of patients between practices and could therefore be used as an instrument for quality improvement. In phase 2, we explored the perceived added value of the indicators in focus group meetings with physical therapists. We then actively involved the participants in the selection of a core set of quality indicators.

Setting

A total of 229 Dutch physical therapists working in 137 primary care practices collected the treatment outcomes of patients with COPD. All participants in the project were recruited via stakeholder organizations in Dutch primary physical therapy care. Participating physical therapists where instructed to treat their patients according to Dutch clinical guideline recommendations for patients with COPD. (16) We only measured outcomes of the treatment; the physical therapists individually decided which treatment was needed for their patients. All procedures were conducted according to the Declaration of Helsinki and approved by the Medical Ethical Committee of Radboud university medical center (Registration # 2019-5455). The STROBE-checklist was used to report the current study. (17) Furthermore, a

framework with tools to support the selection and implementation of patient-reported outcome measures was used as guidance for conduction of this study.(3)

Data collection

Data on the treatment outcomes were anonymously collected through electronical health records (EHRs) via three databases, the national data registry (LDK) of the Association for Quality in Physical Therapy (SKF), the national data registry (LDF) of the Royal Dutch Society for Physical Therapy (KNGF), and the database of Spot On Medics (SOM), which is one of the EHR vendors. The EHRs uploaded to the national registries only contain anonymized data. Furthermore, to ensure the uniformity of the provided data, all data in the registries were collected based on predefined technical specifications.(18) Informed consent was obtained and registered in the EHR from all participating patients included in the current study.

Outcome domains and measures

The outcome domains in the standard set were based on the consensus between stakeholders (patients, physical therapists, policy makers, researchers, and health insurers),(4) and on the recommendations in the Dutch CPG for the physical therapy treatment of patients with COPD.(16) After the development of the standard set, the KNGF published an update for this CPG,(10) in which the suggested outcome domains and associated measures to evaluate physical therapy treatment are in line with the outcome domains from the developed standard set.

The standard set of outcome measures consisted of three mandatory measures for the total population, two conditional measures that depended on the treatment goal, and two exploratory measures that were used as pilot in a small subgroup. In the current study, only the mandatory and conditional measures were used for the development of quality indicators, since the exploratory measure was only used in a small subgroup of practices. The three mandatory measures for all patients with COPD were the 6MWT for measuring physical capacity, the Clinical COPD Questionnaire (CCQ) for measuring health-related quality of life, and the Global Perceived Effect - Dutch Version (GPE-DV) for measuring the perceived effect. The two conditional measures were the hand-held dynamometer (with Microfet™) for measuring quadriceps strength, and the Medical Research Council dyspnea scale (MRC) for measuring dyspnea. All measures were completed pre- and post-treatment to monitor the changes in outcomes over time, except the GPE-DV, which was only measured after the treatment. For a description of each measure and the measurement protocol, see Appendix A. All physical therapists followed a specific protocol to standardize the testing procedure.(19)

Inclusion and exclusion criteria

All patients with COPD (GOLD I–IV), as diagnosed by a medical doctor, who received physical therapy in one of the participating primary care practices between February 2018 and December 2019 were included. Participating physical therapy practices were instructed to collect at least all mandatory and conditional measures from the standard set as presented in the data collection, according to the measurement protocol described in Appendix A. Based on a rule of thumb, a minimum of 30 patients should preferably be included for each practice to allow a valid comparison. (20, 21) However, it was expected that this inclusion requirement could not be reached due to the short inclusion period, and the fact that routine data collection in primary physical therapy care for patients with COPD is relatively new. Therefore, we used a lower threshold and physical therapy practices were excluded from the analysis for a specific quality indicator if they included fewer than 10 patients with COPD

Phase 1: Defining quality indicators and estimating their comparability and discriminability.

Defining potential quality indicators

We used national and international standards for defining potential quality indicators. (2, 5, 22, 23) Quality indicators can be described using mean values and between-relative differences, or quantified and expressed as a proportion in which the numerator describes the number of 'correct' scores and the denominator is the number of persons for which the quality indicator is applicable.(3) See Table 1 for an example of a quality indicator for physical capacity measured with the 6MWT.

Table 1 Example of a quality indicator monitoring the repeated measurement of the 6MWT

Definition	The proportion of patients with COPD who underwent physical therapy treatment and who completed the 6MWT pre- and post-treatment to evaluate physical capacity
Rationale	Improvement of physical capacity is an important goal in physical therapy treatment for patients with COPD. Physical capacity is measured with the 6MWT
Numerator	The number of patients who underwent physical therapy treatment and who completed the 6MWT pre- and post-treatment
Denominator	All patients who underwent physical therapy treatment
Specification	Physical capacity is measured in all patients using the 6MWT, a physical performance test where the patients walk for six minutes in a comfortable way
Type of indicator	Process

For each of the five measures in the standard set, we defined four types of quality indicators: 1) by monitoring the process, i.e., whether the outcome was actually measured pre- and post-treatment; 2) by using mean end scores of the outcome, reflecting patient functioning at the end of treatment; 3) by using the mean pre- to post-treatment change in the outcome score, reflecting improvement or decline in

the outcome; and 4) by using the minimally clinical important difference (MCID) of the outcome, i.e., the proportion of patients who experienced clinically relevant improvements, stabilizations, or deteriorations (see Box 1 for an example). The change score and the MCID were not defined for the GPE-DV, as this measure was only completed after the treatment. For the MRC, no MCID was defined, since a MCID has not yet been established for the MRC.(24) Hence, in total, we defined 17 potential quality indicators. See Appendix B for an extensive description of each potential quality indicator.

Box 1: Potential process and outcome quality indicators at the physical therapist or practice level

a) Process indicator: proportion of repeated measures

The proportion of patients with COPD who underwent physical therapy treatment in which a pre- and post-measurement was used.
Example 1: In 60% of the patients, physical capacity was measured pre- and post-treatment with the 6MWT

b) Outcome indicator: mean end scores

The mean end score (with 95% confidence intervals (CI)) of patients with COPD after a physical therapy treatment.
Example 2 The mean end score of the health-related quality of life of patients with COPD measured with the CCQ is 2.2 points (± 0.9 points)

c) Outcome indicator: mean change scores

The mean change score (with 95% CI) of patients with COPD between the pre- and post-physical therapy stages.
Example 3 The mean change score in the symptoms of dyspnea in patients with COPD measured with the MRC is 2.5 points of improvement (± 1.0 points)

d) Outcome indicator: MCID

The proportion (with 95% CI) of patients with COPD who experienced a MCID improvement between the pre- and post-treatment stages.
Example 4: In 70% (± 7%) of the patients, a clinically relevant change in quadriceps strength was reported after treatment, as measured with the HHD.
Abbreviations: 6MWT: Six-Minute Walk Test; CCQ: Clinical COPD Questionnaire; GPE-DV: Global Perceived Effect - Dutch Version; HHD: hand-held dynamometer; MRC: Medical Research Council dyspnea scale; MCID: minimal clinical important difference; CI: confidence intervals.

Estimating the comparability and discriminability of the quality indicators

For each measure, at least 30 physical therapy practices needed to be included, this was based on a rule of thumb in multilevel analysis for general calculation. (20, 21) Descriptive statistics were used to determine whether the thresholds for

the completeness of the measures were met for estimating the indicator scores. For the analysis, we used measures collected at the beginning and/or end of the treatment. When the treatment episode had not ended, we used the last provided data. The treatment episode is a unique episode of a patient being treated by a physical therapist

Comparability

In a linear and logistic multi-level analysis, patients were clustered within physical therapy practices to compare the outcomes of the quality indicators between practices.(25) The quality indicators were adjusted for patient characteristics that influence the outcome but were not under the control of the physical therapist or physical therapy practice (so called explanatory variables).(23) Explanatory variables, such as age, gender, and baseline scores for each measure, were used for the adjustment of the multi-level analyses. In the analysis, we started with an intercept-only model that estimates only the intercept and the random variation around the intercept. The inclusion of age,(26, 27) gender,(26) and the baseline score (28, 29) of each measure for the adjustment of a multi-level analysis is common in the field of quality indicator development and the comparison of provider performance.(2, 25-30) Next, all explanatory variables were added to the adjusted model, and the influence of the explanatory variables was evaluated by the amount of the random intercept variance that was explained.(25)

For each physical therapy practice in the case-mix adjusted multi-level analysis, the mean scores were estimated with a 95% confidence interval (CI).

Discriminability

For the estimation of the variation in the outcomes between physical therapy practices, the intraclass correlation coefficient (ICC) was calculated. The ICC for physical therapy practices was defined by dividing the variance between practices by the summation of the variance between and the variance within physical therapy practices.(25) In multi-level analyses, most ICCs are between .05 and .20, and ICCs >.10 can be interpreted as adequate, indicating that the quality indicator is able to discriminate outcomes between physical therapists or practices.(25, 31) The ICC was also used to compare the intercept-only model with the adjusted model containing the explanatory variables (case-mix).

Visual representation of indicator scores

To present the mean outcomes for each defined quality indicator of each physical therapy practice in one graph, we used caterpillars plots, as they are found to be user friendly and easy to interpret.(2, 29, 31, 32) We used relative norms by presenting the plots in three colors: blue (95% CI significantly lower than average), purple (no significant 95% CI difference from average), and green (95% CI significantly higher

than average). The plots were used to present the outcomes of the cohort data to the participants in phase 2 of the study.

Phase 2: Selecting a core set of quality indicators

Semi-structured focus group interviews were conducted with purposefully selected participating physical therapists who collected data in phase 1. We also organized one focus group meeting with Dutch senior physical therapists and senior researchers who were members of the development group of the revised Dutch physical therapy guideline for COPD. The senior researchers were asked to comment on the set of quality indicators from a scientific perspective. The senior physical therapists and researchers had at least 10 years of experience in the treatment of and/or research into patients with COPD.

We aimed to conduct four focus group meetings with 6–10 members in every meeting. The primary goal of the focus groups was to reflect on the added value of using the presented indicators in daily practice for quality improvement, and most importantly, to select a core set of quality indicators from the described 17 potential indicators. The focus group meetings where audio recorded and summarized by researcher AV, the summaries of the different focus groups were discussed and interpreted in several meetings with researchers AV (physical therapist and PhD student), SvD (physical therapist and senior researcher), and PvdW (physical therapist and professor of allied health sciences). The identities of the physical therapists were considered confidential; therefore, the answers given by the physical therapists during the interviews and in the survey were processed anonymously. The focus groups were part of the process of reaching consensus on the selection of the core set.

The research members AV, SvD, PvdW, HK (physical therapist and senior researcher), RN (physical therapist and em. professor of allied health sciences) were trained and had experience in conducting qualitative research.

In each focus group meeting, we presented each potential quality indicator in a caterpillar plot, with scores at the levels of the physical therapist and the physical therapy practice, and compared them with the scores for the other practices. The participants interpreted the comparability and discriminability of the potential quality indicators. Finally, we asked participants to select their preferred quality indicators for the core set from the potential quality indicators as described in box 1. During each meeting, the chairman (AV) summed up all the preferred quality indicators and asked the group whether they accepted or declined the proposed core set. A consensus was reached if >80% of the participants accepted the selection of each quality indicator in the core set.

Patient and public involvement

For the development of this standard set we interviewed patients about their perspectives on the selection of patient outcomes. *(4, 15)* Furthermore, during the conduction of this study a steering committee with representatives from important stakeholders, including the association for patients with COPD Netherlands Patients Federation (Longfonds), advised during the selection process. During the meetings, we discussed the views and perspectives of stakeholders regarding the value and implementation of outcome-based quality indicators for Dutch physical therapy

RESULTS

Phase 1: Estimating the comparability and discriminability of the quality indicators

Descriptive statistics

Table 2 shows descriptive statistics of the included treatment episodes and the number of physical therapists and physical therapy practices who provided the data. The treatment episode is a unique episode of a patient being treated by a physical therapist. The current national data registries cannot detect recurrences of patients over time due to privacy regulations; therefore, the number of unique patients may be lower. Overall, 229 physical therapists from 137 practices provided 4651 treatment episodes of patients with COPD.

Table 2 Descriptive statistics of the included patients and the number of participating physical therapists and physical therapy practices

Number of treatment episodes in the dataset	4651
Female patients	2440 (52.5%)
Age, years	67.9 (9.4)
Treatment sessions	49.2 (58.2)
Episode duration, weeks	46.6 (50.3)
Physical therapists who provided the data	229
Physical therapy practices that provided the data	137

Data are presented as means (standard deviation (SD)) or numbers and percentages of the total population.

Table 3 presents the characteristics and unadjusted outcomes of patients with COPD on each measure of the standard set. The number of patients with end scores differed between the measures. Each measure reached the threshold of at least 30 included physical therapy practices that provided ≥10 cases, except for the HHD, for which only 10 physical therapy practices provided ≥10 cases and therefore no ICC was calculated. See Table 3 for the number of practices and provided cases that were included in the multi-level analysis.

Table 3 Descriptive characteristics and unadjusted outcomes of patients with COPD for each measure of the standard set

	6MWT	CCQ	GPE-DV	HHD	MRC
Female patients	1344 (51.1%)	1786 (52.1%)	636 (50.6%)	223 (51.3%)	1237 (52.7%)
Age, years	67.8 (9.2)	68.1 (9.4)	68.2 (9.4)	68.0 (9.2)	68.2 (9.3)
Treatment episodes with baseline scores	2628	3427	N.A.	435	2348
Treatment episodes with end scores	1822	2408	1256	218	1385
Range of scores on each measure	6–780 m	0–6 points	1–7 points	131–542 Nm	1–5 points
Baseline scores[a]	370.8 m (126.3 m)	2.4 points (0.9)	N.A.	284.2 Nm (96.4 Nm)	3.0 points (1.1)
End scores	373.5 m (130.3 m)	2.2 points (0.9)	3.5 points (1.1)	298.4 Nm (95.5)	3.0 points (1.1)
Change (T_{end}–T_0)	2.7 (86.8)	–0.1 (0.8)	N.A.*	8.4 (50.9)	0.2 (1.3)
MCID improvement	533 (28.7%)[b]	818 (34%)[c]	N.A.*	99 (45.4%)[d]	N.A.**
MCID stabilization	807 (44.3%)[b]	1052 (43.7%)[c]	N.A.*	43 (19.9%)[d]	N.A.**
MCID deterioration	490 (26.9%)[b]	537 (22.3%)[c]	N.A.*	76 (34.7%)[d]	N.A.**
Physical therapists who provided data	145 (63.3%)	202 (88.2%)	117 (51.0%)	46 (20.0%)	168 (46.7%)
Practices that provided data	86 (62.8%)	126 (92%)	72 (52.6%)	28 (20.4%)	107 (78.1%)
Practices that provided ≥10 cases	44 (19.2%)	61 (26.6%)	35 (15.3%)	10 (4.4%)	43 (18.8%)
Patients included in the multi-level analysis	1679 (36.0%)	2201 (47.3%)	1110 (23.8%)	160 (3.4%)	1226 (26.4%)

Data are presented as means (standard deviation (SD)) or numbers and percentages of patients with baseline measures

6MWT: Six-Minute Walk Test; CCQ: Clinical COPD Questionnaire; GPE-DV: Global Perceived Effect - Dutch Version; HHD: hand-held dynamometer; MRC: Medical Research Council dyspnea scale; MCID: minimal clinical important difference; CI: confidence intervals; N.A: not applicable.

*GPE-DV was only analyzed at the end of the treatment

** The MCID for the MRC is yet to be established (24)

[a] For treatment episodes with end scores

[b] For the multi-level analysis of the 6MWT, we used an MCID of ± ≥30 m (33)

[c] For the multi-level analysis of the CCQ, we used an MCID of ± ≥0.4 points (34)

[d] For the multi-level analysis of the HHD, we used an MCID of ± ≥7.5 Nm (35)

Comparability

Table 4 presents the ICC calculations of the intercept-only models and the models adjusted with the explanatory variables. For the process measures, similar to the HHD, no ICC was calculated due to insufficient data. In total, 11 models could be estimated for the outcome indicators based on scores of the 6MWT (four indicators), CCQ (four indicators), GPE (one indicator for the end score), and MRC (two indicators; MCID could not be calculated). In 10 of the 11 case-mix adjusted models, the ICC increased compared with the intercept-only model, thus improving the comparability between practices, i.e., the random intercept variance of physical therapy practices increased in the adjusted models

Table 4 ICCs for the intercept-only model and adjusted model for the change, end, and MCID scores for each measure of the total population in practices that provided ≥10 patients

	Intercept-only model	**Adjusted model**
6MWT end score	0.08	0.17[a]
6MWT change score	0.00	0.01[a]
6MWT MCID improvement[b]	0.03	0.04[a]
6MWT MCID deterioration[b]	0.06	0.06
CCQ end score	0.11	0.20[a]
CCQ change score	0.06	0.09[a]
CCQ MCID improvement[c]	0.05	0.07[a]
CCQ MCID deterioration[c]	0.03	0.05[a]
GPE-DV end score	0.14	0.15[a]
MRC end score	0.08	0.12[a]
MRC change score	0.23	0.34[a]

ICC: intraclass correlation coefficient; 6MWT: Six-Minute Walk Test; CCQ: Clinical COPD Questionnaire; GPE-DV: Global Perceived Effect - Dutch Version; HHD: hand-held dynamometer; MRC: Medical Research Council dyspnea scale; MCID: minimal clinical important difference; CI: confidence intervals; N.A: not applicable.

[a] Increase in the ICC compared with the intercept-only model following the adjustment for the case-mix variables age, gender, and baseline score of the measure.

[b] For the multi-level analysis of the 6MWT, we used an MCID of ± ≥30 m (33)

[c] For the multi-level analysis of the CCQ, we used an MCID of ± ≥0.4 points (34)

Discriminability

Five of the 11 case-mix adjusted ICCs were >0.10, ranging between 0.12 and 0.32, which can be interpreted as an adequate discriminability. All adjusted models were used for the visual representation of the quality indicators in the focus group interviews. Also, the quality indicators presenting the outcomes of the HHD, for which no multi-level analysis was conducted, were presented in the focus group interviews. All defined potential quality indicators are presented as caterpillar plots

(figures 2a-c). Each graph shows that a wide range of differences in outcomes exist between physical therapy practices.

Phase 2: Selecting a core set of quality indicators

In total, four focus group interviews were conducted with 20 (out of 22 invited) physical therapists and three (out of five invited) senior researchers. The non-acceptance of invited participants was due to the date and time of the focus groups that did not fit with the agenda of the potential participants. The mean duration of the focus groups was 90 minutes (range 80 – 95) Nine were female, the mean age of the participants was 39 years (range 23–60 years), and they had an average work experience of 14 years (range 1–35 years). In total, 16 of the 20 participating physical therapists also provided data for the practice test. See appendix C for an overview of the characteristics of the participants. Almost all the participants expressed that the presented quality indicators were user friendly and had value for quality improvement in daily practice. Still, several issues surrounding the presented quality indicators were discussed.

Using patient profiles for the comparison of patient outcomes

The participants mentioned that, in future research, it would be helpful to stratify patients based on the Dutch model, a profiling system to enhance the comparability between physical therapy practices. In 2020, an ad hoc task force of experts in the field of physical therapy, exercise therapy, rehabilitation science, respiratory medicine, general medicine, and elderly care medicine, as well as patient representatives, developed a profiling system (the "Dutch model") for patients with COPD to allocate patients into subgroups for exercise-based care.(36) The participants of the current study suggested that baseline measures and patient characteristics needed for allocating patients to subgroups according to the profiling system should be included as process indicators in the core set. They stated that the stratification of patients into subgroups based on these profiles would enhance the comparability between practices.

Conditions for interpreting outcomes

Another reported problem was the limited amount of provided outcome data in the study, especially for the HHD. A possible reason could be that for the data collection we used real-world data via national data registries. These registries used predefined technical specifications. (18) During the conduction of the study, the HHD was a new measure implemented in the data registries. Potentially, this may have resulted in the low amount of provided data, which was also mentioned in the focus groups. Participant therefore suggested that the implementation of process measures is needed to stimulate routine data collection as a first step in quality improvement. When comparing outcomes, the participants were interested in the background information of patients with COPD, such as smoking status, exacerbations, and body

weight, for better interpretation of the differences in, for example, the change or end scores. When using these outcomes as a learning tool, the education of physical therapists is needed to gain knowledge about the interpretation of the outcomes. Furthermore, to enhance the comparability between practices, participants suggested to include only outcomes of patients that were treated for ≥3 months.

Including the percentage of a predicted value

Absolute outcomes were used to calculate the end and change scores for the 6MWT and HHD. The participants suggested that outcomes should be presented as percentages of predicted values based on reference data from the healthy population.(37, 38) These normative values are based on previous research and can be calculated according to gender, age, and body weight.

Selection of the core set

After discussing the outcomes, all (100%) of the physical therapists in each focus group preferred the inclusion of seven quality indicators in the core set: three process indicators for the routine measurement of the 6MWT, CCQ, and HHD; three outcome indicators using the pre- to post-treatment change in the 6MWT, CCQ, HHD scores; and a combined process indicator to monitor the baseline measurement of three measures (6MWT, CCQ, and an accelerometer (steps per day)) and patient characteristics (age, gender, body weight, and number of exacerbations in the past year) to allocate patients into subgroups based on the profiling system of the Dutch model.(36) The final core set of seven quality indicators is shown in Table 5. Figures 1a (6MWT) 1b(CCQ) 1c(HHD) presents the proportion of patients with pre- and post-treatment and 2a (6MWT), 2b (CCQ) and 2c (HHD) presents the caterpillar plots of the quality indicators in the final core set.

Table 5 Selected core set of quality indicators accepted by stakeholders based on the perceived added value as quality improvement tools

Type of indicator	Quality indicator description	Overall mean/ Percentage*	Range*
Physical capacity measured with the 6MWT			
Process	The proportion of patients with COPD who underwent physical therapy treatment and who completed the 6MWT pre- and post-treatment to evaluate physical capacity	60.7%	26.1–88.8%
Outcome	The mean change score ± 95% CI of patients with COPD who underwent physical therapy treatment and pre- and post-treatment measurement with the 6MWT to evaluate physical capacity	2.8 meters	–5.4 – 13.4
Health-related quality of life measured with the CCQ			
Process	The proportion of patients with COPD who underwent physical therapy treatment and who completed the CCQ pre- and post-treatment to evaluate aspects of health-related quality of life	62.6%	14.8–88.7%
Outcome	The mean change score ± 95% CI of patients with COPD who underwent physical therapy treatment and pre- and post-treatment measurement with the CCQ to evaluate health-related quality of life	–0.1	0.3 – –0.6
Quadriceps strength measured with the HDD			
Process	The proportion of patients with COPD who underwent physical therapy treatment and who completed the HHD pre- and post-treatment to evaluate quadriceps strength	31.4%	5.9–87.5%
Outcome	The mean change score ± 95% CI of patients with COPD who underwent physical therapy treatment and pre- and post-treatment measurement with the HHD to evaluate quadriceps strength	7.5 Nm	2.7–13.1
Baseline measures for the 6MWT, CCQ, accelerometer, and patient characteristics that can be used in a profiling system to stratify patients into subgroups for care**			
Process	The proportion of patients with COPD who underwent physical therapy treatment and who completed the baseline measurements for the 6MWT, CCQ, accelerometer, gender, age, body weight, and number of exacerbations in the past year	2.4%	

6MWT: Six-Minute Walk Test; CCQ: Clinical COPD Questionnaire; GPE-DV: Global Perceived Effect - Dutch Version; HHD: hand-held dynamometer; MRC: Medical Research Council dyspnea scale; MCID: minimal clinical important difference; CI: confidence intervals. Only outcomes of patients that were included that were treated for ≥3 months

* The overall mean/percentage and range are the outcomes of the physical therapy practices that provided ≥10 cases, used for describing and selecting the quality indicators for the core set

** Baseline measures and patient characteristics selected to allocate patients into subgroups based on the Dutch model for exercise-based care in primary care (36)

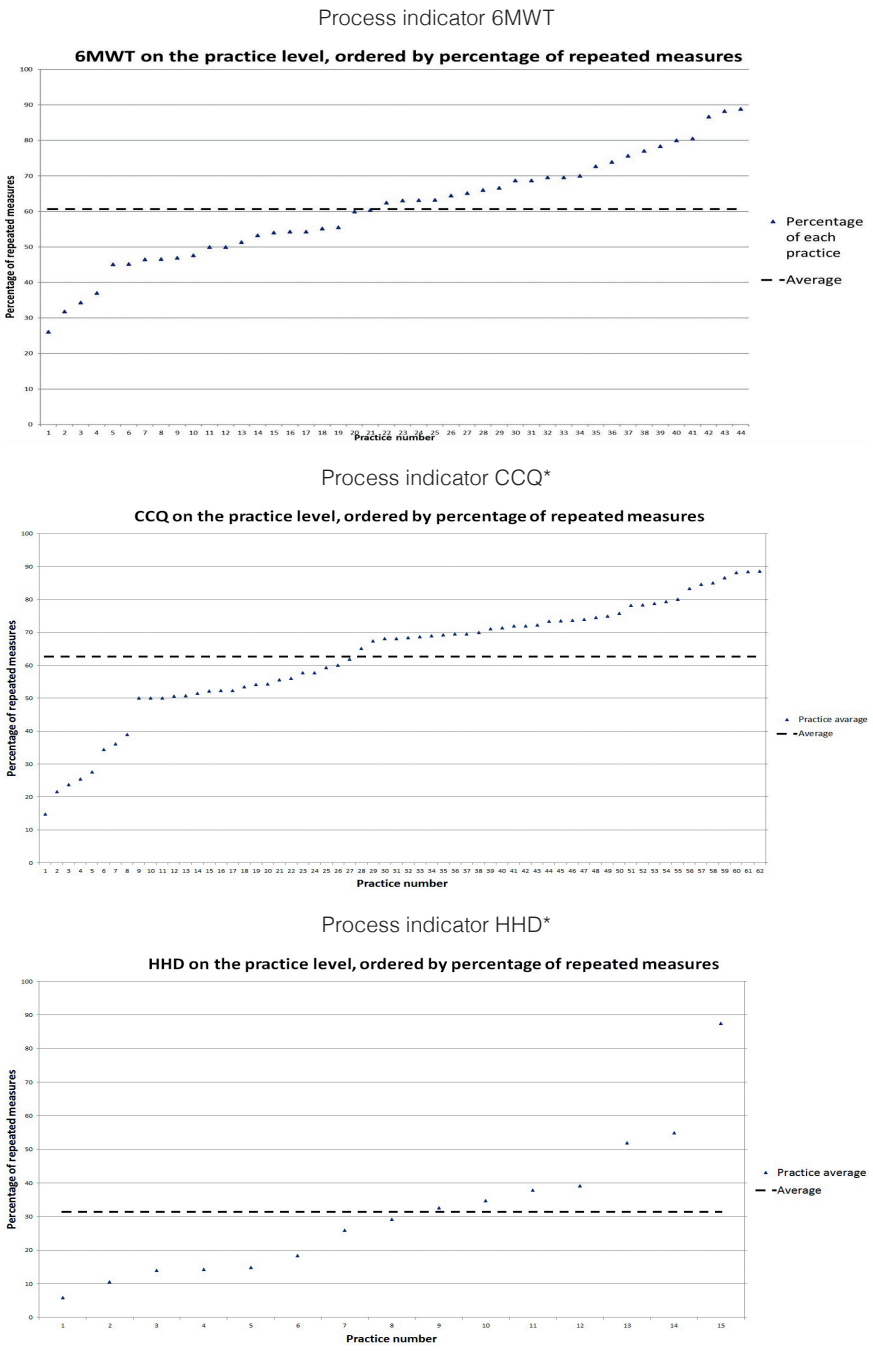
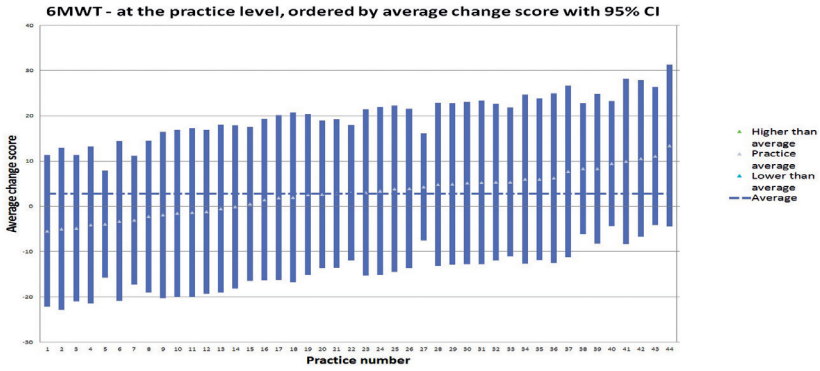


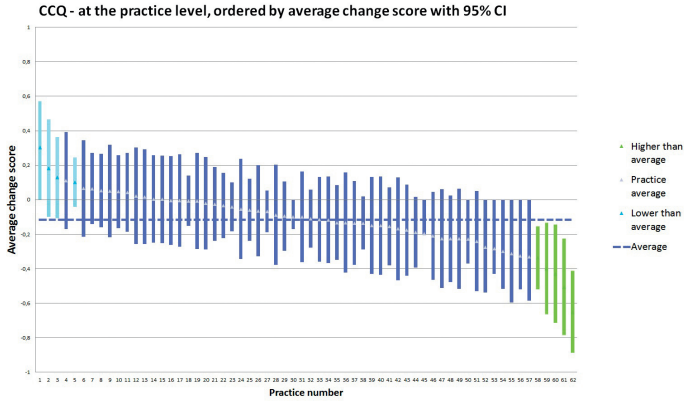
Figure 1: Visual representation of process quality indicators on practice level

* The proportion of patients with COPD who underwent physical therapy treatment in which a pre and/or post the measure was provided

Change score 6MWT*



Change score CCQ*



Change score HHD*

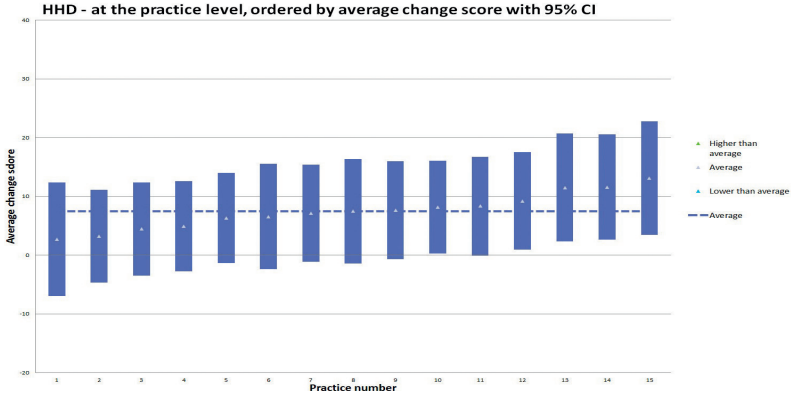


Figure 2: Visual representation outcome quality indicators on practice level

* The mean change score with 95% CI of patients with COPD between pre- and post-physical therapy treatment

DISCUSSION

The major finding in this study is that all participants in the focus groups accepted the quality indicators as a quality improvement tool based on their perceived added value, and selected a core set of seven outcome-based quality indicators for patients with COPD. The final core set includes a process and outcome indicator for three outcomes: physical capacity measured with the 6MWT, health-related quality of life measured with the CCQ, and quadriceps strength measured with the HHD. A combined process indicator was included to monitor the baseline measurement of three measures used to allocate patients into subgroups based on the Dutch model profiling system.(36) To our knowledge, this is the first study to develop a core set of outcome-based quality indicators including a practice test for patients with COPD in physical therapy primary care practice. With the use of the core set, it is possible to compare standardized outcomes for patients between practices.

Several studies have developed quality indicators for COPD care,(39-44) but most sets were developed for the evaluation of processes or structures of care, e.g., monitoring the proportion of patients for whom smoking status was recorded or the availability of exercise equipment.(39, 41-44) These studies differed in their care focus areas, which were hospitalized care, end-of life care, transitional care after hospitalization, pulmonary rehabilitation, vulnerable elders, or primary care in general. (39-44) None of these publications performed a practice test. In one indicator set, developed for pulmonary rehabilitation, some similar domains (physical capacity, strength, and health-related quality of life) and measures (6MWT) were described. (43) The selection of change scores in the core set and the use of caterpillar plots is in line with other research describing the development of quality indicators based on PROMs. (2, 5, 29) (2, 5, 29) A difference is that in the current study we focused in specific on development quality indicators based on outcomes of care, while other studies are more focused on evaluating processes of care. (7, 45-49) None of these studies aimed to develop a core set of outcome-based quality indicators to be used as quality improvement tools for healthcare providers, however. Quality indicators can also be developed for pay-for-performance initiatives, policy reports, insight into practice variation/delivered care, or the identification of differences in delivered care.

Despite the fact that our core set was developed in a Dutch environment, physical therapists in other countries could potentially use the indicator set in their daily practice. Nonetheless, the context of each country needs to be taken into account, specifically cultural or clinical practice differences between countries, such as differing guidelines or educational levels of physical therapists.(50)

A strength of the current study is that we used a standard set of outcome domains and associated measures. The standard set was explicitly developed for patients

with COPD being treated in primary care physical therapy practices, which was designed to be used as a basis for the further development of quality indicators.(4) The standard set is based on recommendations in guidelines and the supporting literature, and was selected in a RAND/UCLA Delphi procedure, which is one of the most common methods for the development of quality indicators.(11, 51)

Another strength of our study is that we collected real-world data to perform a practice test prior to the selection of the core set, which was judged as an essential step in evaluating the validity, reliability, and feasibility of the indicators.(9) The interpretation of the practice test was discussed with end-users and guideline developers in focus groups. Including stakeholders in the development process is an important step for the successful implementation of quality indicators.(51) In the current study, we explicitly focused on the development of an indicator set for learning and quality improvement purposes for physical therapists. When quality indicators are designed for other purposes, such as a support tool for patients to choose providers, future research should also include other stakeholders (i.e., patients, policy makers, and health insurers) for the evaluation of their usefulness in daily practice.

A limitation of our study is that in multi-level analyses, a general rule of the thumb for the calculation of outcomes is the 30/30 rule (i.e., 30 physical therapy practices including a minimum of 30 patients each), allowing a valid comparison of indicator scores between practices.(20, 21) We did not use this rule of thumb as the threshold for estimating the case-mix adjusted scores for each quality indicator. The routine collection of clinical data by Dutch physical therapists treating patients with COPD is still in its infancy; therefore, we concluded that the 30/30 rule would not have been achievable in our study. Here, the collected data was only used as supporting tool for the selection of the core set, so we decided to include physical therapy practices that had ≥10 patients with COPD.

It is important to note that many practices did not reach the threshold of providing measurements for ≥10 patients with COPD. When the process indicators, as presented in Table 5, were based on all participating practices, the proportion of repeated measures was 39% for the 6MWT, 52% for the CCQ, and 5% for the HHD. In our view, future implementation strategies must be conducted to improve the amount of data provided; for example, by giving feedback to practices with process indicators as presented in Table 5.

Furthermore, due to the amount of data provided, we chose to compare the outcomes between physical therapy practices and not between physical therapists. When the amount of available data increases, the opportunity to compare outcomes between physical therapists, both between and within practices, will arise. When sufficient data within practices is provided, physical therapists are able to learn from their own outcomes in comparison with peers who are employed in the same practice. We expect that, when comparing outcomes between physical therapists, the variability will be larger than between practices.

Another limitation is that we were not able to collect data that allowed us to allocate patients into subgroups based on their burden of disease, physical activity, and physical capacity.(4) Hypothetically, the comparability and discriminability of the quality indicators would increase when allocating patients into subgroups. The participants of the focus groups underlined this hypothesis and suggested the inclusion of the Dutch profiling system for patients with COPD in the core set;(36) however, the Dutch profiling system had not yet been developed at the start of the data collection for this study. Future research could evaluate the core set for each subgroup to compare more homogeneous patient groups on their baseline characteristics. Another aspect to increase the comparability is to include more patient characteristics for case-mix adjustment. As suggested by patients with COPD and physical therapists, potential relevant case-mix variables are, for example, smoking history, comorbidities and number of exacerbations.(4)

Implications for practice

Outcome-based quality indicators based on real-world data, as provided in this study, can be used as a learning tool by comparing the collected patient outcomes between physical therapists or practices. This can, for example, be accomplished by discussing outcomes in peer assessment meetings of physical therapists to improve the quality of care. In such meetings, physical therapists critically appraise their peers' performance and give them constructive feedback. *(52-54)* In our opinion, Dutch physical therapists treating patients with COPD should first focus on expanding the amount of data collected. Giving feedback information can help to stimulate physical therapy practices to increase data collection. When sufficient data is provided and the comparison of outcomes in patient subgroups is established, the usability of the core set will increase. Future research should focus on the development of methods to improve the use of outcomes between peers and to set up specific actions to improve the quality of care.

Conclusion

This is the first study to describe and select a core set of seven outcome-based quality indicators for patients with COPD treated in primary care physical therapy practice. This core set includes process and outcome indicators related to measuring physical capacity, health-related quality of life, and quadriceps strength, and a process measure for profiling patients within subgroups. To further evaluate the core outcome set, future research should explore different strategies to promote data collection, including providing feedback of the outcomes to physical therapists.

Acknowledgments: We acknowledge our research members Juliette Cruijsberg and Janine Liefers; senior researcher Harald Miedema; senior physical therapists Cor Zagers and Leendert Tissink; and all the physical therapists who participated in this study and provided data and/or participated in the focus group meetings.

REFERENCES

1. Black N. Patient reported outcome measures could help transform healthcare. Bmj. 2013;346:f167.
2. Hildon Z, Neuburger J, Allwood D, van der Meulen J, Black N. Clinicians' and patients' views of metrics of change derived from patient reported outcome measures (PROMs) for comparing providers' performance of surgery. BMC health services research. 2012;12:171.
3. van der Wees PJ, Verkerk EW, Verbiest MEA, Zuidgeest M, Bakker C, Braspenning J, et al. Development of a framework with tools to support the selection and implementation of patient-reported outcome measures. Journal of Patient-Reported Outcomes. 2019;3(1):75.
4. Verburg AC, van Dulmen SA, Kiers H, Ypinga JH, Nijhuis-van der Sanden MW, van der Wees PJ. Development of a Standard Set of Outcome Domains and Proposed Measures for Chronic Obstructive Pulmonary Disease in Primary Care Physical Therapy Practice in the Netherlands: a Modified RAND/UCLA Appropriateness Method. International journal of chronic obstructive pulmonary disease. 2019;14:2649-61.
5. Patient Reported Outcomes (PROs) in Performance Measurement Washington, DC: National Quality Form (NQF); 2012 [Available from: http://www.qualityforum.org/Publications/2012/12/Patient-Reported_Outcomes_in_Performance_Measurement.aspx.
6. Jette DU, Halbert J, Iverson C, Miceli E, Shah P. Use of standardized outcome measures in physical therapist practice: perceptions and applications. Phys Ther. 2009;89(2):125-35.
7. Westby MD, Marshall DA, Jones CA. Development of quality indicators for hip and knee arthroplasty rehabilitation. Osteoarthritis and cartilage. 2018;26(3):370-82.
8. Fitch K, Steven J. Bernstein, Maria Dolores Aguilar, Bernard Burnand, Juan Ramon LaCalle, Pablo Lazaro, Mirjam van het Loo, Joseph McDonnell, Janneke Vader and James P. Kahan. The RAND/UCLA Appropriateness Method User's Manual. RAND Corporation. 2001.
9. Kötter T, Blozik E, Scherer M. Methods for the guideline-based development of quality indicators--a systematic review. Implementation Science. 2012;7(1):21.
10. Vreeken HL, Beekman E, Doormaal van MCM, Post MHT, Meerhoff GA, Spruit M. Royal Dutch Society for Physical Therapy (KNGF) practice guideline COPD 2020.
11. Langendam MW, Piggott T, Nothacker M, Agarwal A, Armstrong D, Baldeh T, et al. Approaches of integrating the development of guidelines and quality indicators: a systematic review. BMC health services research. 2020;20(1):875.
12. Wollersheim H, Hermens R, Hulscher M, Braspenning J, Ouwens M, Schouten J, et al. Clinical indicators: development and applications. Neth J Med. 2007;65(1):15-22.
13. Meerhoff GA, Verburg AC, Schapendonk RM, Cruijsberg J, Nijhuis-van der Sanden MWG, van Dulmen SA, et al. Reliability, validity and discriminability of patient reported outcomes for non-specific low back pain in a nationwide physical therapy registry: A retrospective observational cohort study. PLoS One. 2021;16(6):e0251892.
14. Verburg AC, van Dulmen SA, Kiers H, Nijhuis-van der Sanden MWG, van der Wees PJ. Patient-Reported Outcome-Based Quality Indicators in Dutch Primary Care Physical Therapy for Patients With Nonspecific Low Back Pain: A Cohort Study. Phys Ther. 2021.
15. Core Outcome Measures in Effectiveness Trials (COMET) Initiative 2019 [Available from: https://www.comet-initiative.org/Studies/Details/1469.

16. Gosselink R, Langer D, Burtin C, Probst V, Hendriks HJM, Schans CPvd, et al. KNGF-Guideline for physical therapy in patients with chronic obstructive pulmonary diseasehttps://www.fysionet-evidencebased.nl/images/pdfs/richtlijnen/copd_2008/copd_verantwoording_en_toelichting.pdf. 2008.

17. Cuschieri S. The STROBE guidelines. Saudi J Anaesth. 2019;13(Suppl 1):S31-s4.

18. Meerhoff GA, van Dulmen SA, Maas MJM, Heijblom K, Nijhuis-van der Sanden MWG, Van der Wees PJ. Development and Evaluation of an Implementation Strategy for Collecting Data in a National Registry and the Use of Patient-Reported Outcome Measures in Physical Therapist Practices: Quality Improvement Study. Physical Therapy. 2017;97(8):837-51.

19. Guideline protocol to standardize the testing procedure for patients with COPD (in Dutch). 2018.

20. Ali A, Ali S, Khan SA, Khan DM, Abbas K, Khalil A, et al. Sample size issues in multilevel logistic regression models. PloS one. 2019;14(11):e0225427-e.

21. Koolman X, Zuidgeest M, Visser J, Appelman M. Indicatorstandaard 2.0: Methodologische criteria voor de ontwikkeling van betrouwbare kwaliteitsindicatoren in de zorg. : Kwaliteitsinstituut; 2012.

22. Kessel Pv, Triemstra, M., Boer, D. de. Handreiking voor het meten van kwaliteit van zorg met Patient Reported Outcome Measures 2014 [Available from: https://www.nivel.nl/sites/default/files/bestanden/Handreiking-meten-kwaliteit-zorg-met-Patient-Reported-Outcome-Measures.pdf.

23. Greenhalgh J, Dalkin S, Gooding K, Gibbons E, Wright J, Meads D, et al. Health Services and Delivery Research. Functionality and feedback: a realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care. Southampton (UK): NIHR Journals Library; 2017.

24. Cazzola M, Hanania NA, MacNee W, Rüdell K, Hackford C, Tamimi N. A review of the most common patient-reported outcomes in COPD--revisiting current knowledge and estimating future challenges. International journal of chronic obstructive pulmonary disease. 2015;10:725-38.

25. Twisk JWR. Applied Mixed Model Analysis: A Practical Guide. 2 ed. Cambridge: Cambridge University Press; 2019.

26. Gray J, Millett C, O'Sullivan C, Omar RZ, Majeed A. Association of age, sex and deprivation with quality indicators for diabetes: population-based cross sectional survey in primary care. J R Soc Med. 2006;99(11):576-81.

27. Kiderlen M, Ponti A, Tomatis M, Boelens PG, Bastiaannet E, Wilson R, et al. Variations in compliance to quality indicators by age for 41,871 breast cancer patients across Europe: a European Society of Breast Cancer Specialists database analysis. Eur J Cancer. 2015;51(10):1221-30.

28. Nunes EV, Pavlicova M, Hu M-C, Campbell AN, Miele G, Hien D, et al. Baseline matters: the importance of covariation for baseline severity in the analysis of clinical trials. Am J Drug Alcohol Abuse. 2011;37(5):446-52.

29. Hildon Z, Allwood D, Black N. Making data more meaningful: patients' views of the format and content of quality indicators comparing health care providers. Patient education and counseling. 2012;88(2):298-304.

30. Hildon Z, Allwood D, Black N. Patients' and clinicians' views of comparing the performance of providers of surgery: a qualitative study. Health Expect. 2015;18(3):366-78.

31. Boer D, Triemstra M, Koopman L, Zuidgeest M. Tool Analyses ; voor het vergelijken van patiëntervaringen tussen zorgaanbieders | versie 1.0 2016 [Available from: https://www.zorginzicht.nl/binaries/content/documents/zorginzicht/ontwikkeltools/prom-toolbox/prom-cyclus-stap-5-definieren-indicator/prom-cyclus-stap-5-definieren-indicator/files/3/Tool-Analyse-voor-het-vergelijken-van-pati%C3%ABntervaringen-tussen-zorgaanbieders.pdf.

32. Goldstein H, Michael JRH. The Graphical Presentation of a Collection of Means. Journal of the Royal Statistical Society Series A (Statistics in Society). 1995;158(1):175-7.

33. Bohannon RW, Crouch R. Minimal clinically important difference for change in 6-minute walk test distance of adults with pathology: a systematic review. J Eval Clin Pract. 2017;23(2):377-81.

34. Vaidya T, Beaumont M, de Bisschop C, Bazerque L, Le Blanc C, Vincent A, et al. Determining the minimally important difference in quadriceps strength in individuals with COPD using a fixed dynamometer. International journal of chronic obstructive pulmonary disease. 2018;13:2685-93.

35. Smid DE, Franssen FM, Houben-Wilke S, Vanfleteren LE, Janssen DJ, Wouters EF, et al. Responsiveness and MCID Estimates for CAT, CCQ, and HADS in Patients With COPD Undergoing Pulmonary Rehabilitation: A Prospective Analysis. J Am Med Dir Assoc. 2017;18(1):53-8.

36. Spruit MA, Van't Hul A, Vreeken HL, Beekman E, Post MHT, Meerhoff GA, et al. Profiling of Patients with COPD for Adequate Referral to Exercise-Based Care: The Dutch Model. Sports Medicine. 2020;50(8):1421-9.

37. Troosters T, Gosselink R, Decramer M. Six minute walking distance in healthy elderly subjects. The European respiratory journal. 1999;14(2):270-4.

38. Andrews AW, Thomas MW, Bohannon RW. Normative values for isometric muscle force measurements obtained with hand-held dynamometers. Phys Ther. 1996;76(3):248-59.

39. De Schreye R, Houttekier D, Deliens L, Cohen J. Developing indicators of appropriate and inappropriate end-of-life care in people with Alzheimer's disease, cancer or chronic obstructive pulmonary disease for population-level administrative databases: A RAND/UCLA appropriateness study. Palliat Med. 2017;31(10):932-45.

40. Michael F, Whitelaw S, Van Spall HG. Transitional care quality indicators to assess quality of care following hospitalisation for chronic obstructive pulmonary disease and heart failure: a systematic review protocol. BMJ open. 2019;9(12):e032764.

41. Weidinger P, Nilsson JL, Lindblad U. Adherence to diagnostic guidelines and quality indicators in asthma and COPD in Swedish primary care. Pharmacoepidemiol Drug Saf. 2009;18(5):393-400.

42. Gershon AS, Mecredy GC, Aaron SD, Camp PG, Tu K, Hernandez P, et al. Development of quality indicators for chronic obstructive pulmonary disease (COPD): A modified RAND appropriateness method. Canadian Journal of Respiratory, Critical Care, and Sleep Medicine. 2019;3(1):30-8.

43. Dechman G, Cheung W, Ryerson CJ, Hernandez P, Stickland M, Gershon A, et al. Quality indicators for pulmonary rehabilitation programs in Canada: A Canadian Thoracic Society expert working group report. Canadian Journal of Respiratory, Critical Care, and Sleep Medicine. 2019;3(4):199-209.

44. Kleerup E. Quality indicators for the care of chronic obstructive pulmonary disease in vulnerable elders. J Am Geriatr Soc. 2007;55 Suppl 2:S270-6.

45. Petersson IF, Strombeck B, Andersen L, Cimmino M, Greiff R, Loza E, et al. Development of healthcare quality indicators for rheumatoid arthritis in Europe: the eumusc.net project. Ann Rheum Dis. 2014;73(5):906-8.

46. Peter WF, Hurkmans EJ, Wees PJvd, Hendriks EJM, Bodegom-Vos Lv, Vlieland TPMV. Healthcare Quality Indicators for Physiotherapy Management in Hip and Knee Osteoarthritis and Rheumatoid Arthritis: A Delphi Study. Musculoskeletal Care. 2016;14(4):219-32.

47. Scholte M, Neeleman-van der Steen CW, Hendriks EJ, Nijhuis-van der Sanden MW, Braspenning J. Evaluating quality indicators for physical therapy in primary care. International journal for quality in health care : journal of the International Society for Quality in Health Care. 2014;26(3):261-70.

48. Oostendorp RA, Rutten GM, Dommerholt J, Nijhuis-van der Sanden MW, Harting J. Guideline-based development and practice test of quality indicators for physiotherapy care in patients with neck pain. J Eval Clin Pract. 2013;19(6):1044-53.

49. Oostendorp RAB, Elvers JWH, van Trijffel E, Rutten GM, Scholten-Peeters GGM, Heijmans M, et al. Relationships Between Context, Process, and Outcome Indicators to Assess Quality of Physiotherapy Care in Patients with Whiplash-Associated Disorders: Applying Donabedian's Model of Care. Patient Prefer Adherence. 2020;14:425-42.

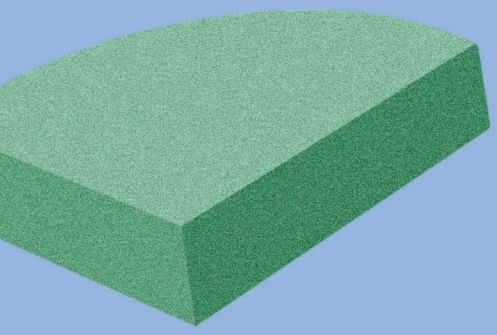
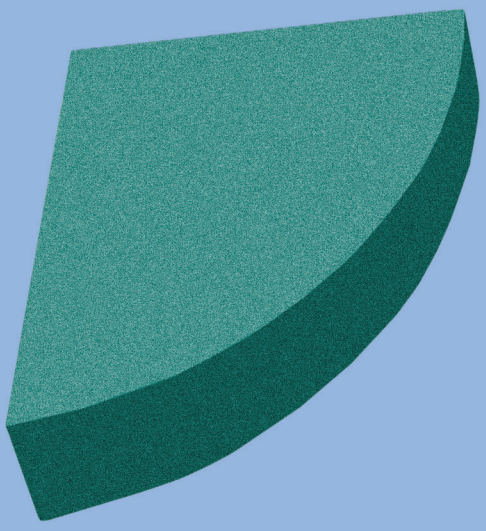
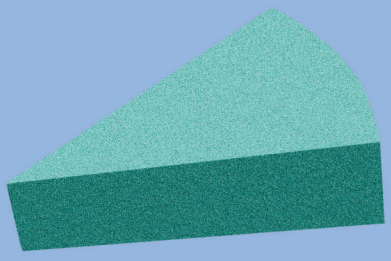
50. Marshall MN, Shekelle PG, McGlynn EA, Campbell S, Brook RH, Roland MO. Can health care quality indicators be transferred between countries? Quality & safety in health care. 2003;12(1):8-12.

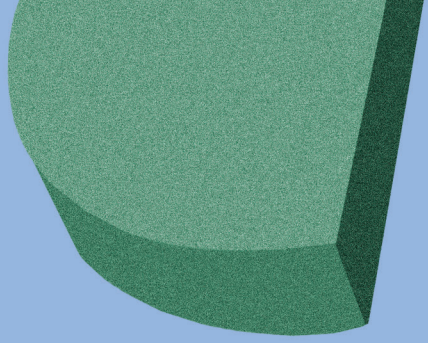
51. Campbell SM, Braspenning J, Hutchinson A, Marshall MN. Research methods used in developing and applying quality indicators in primary care. BMJ. 2003;326(7393):816-9.

52. Maas MJ, Nijhuis-van der Sanden MW, Driehuis F, Heerkens YF, van der Vleuten CP, van der Wees PJ. Feasibility of peer assessment and clinical audit to self-regulate the quality of physiotherapy services: a mixed methods study. BMJ open. 2017;7(2):e013726.

53. Maas MJM, Driehuis F, Meerhoff GA, Heerkens YF, van der Vleuten CPM, Nijhuis-van der Sanden MWG, et al. Impact of Self- and Peer Assessment on the Clinical Performance of Physiotherapists in Primary Care: A Cohort Study. Physiother Can. 2018;70(4):393-401.

54. Maas MJ, van der Wees PJ, Braam C, Koetsenruijter J, Heerkens YF, van der Vleuten CP, et al. An innovative peer assessment approach to enhance guideline adherence in physical therapy: single-masked, cluster-randomized controlled trial. Phys Ther. 2015;95(4):600-12.



CHAPTER 6

Experiences of physiotherapists regarding a standard set of measurement instruments to improve quality of care for patients with chronic obstructive pulmonary disease

A mixed methods study

Arie C. Verburg
Jessica Zincken
Simone A. van Dulmen
Henri Kiers
Philip J. van der Wees

Accepted for publication in Journal of Patient-Reported Outcomes

ABSTRACT

Rationale: The quality of physiotherapy care for patients with chronic obstructive pulmonary disease (COPD) can be improved by comparing outcomes of care in practice.

Aim: To evaluate the experiences of physiotherapists implementing a standard set of measurement instruments to measure outcomes and improve the quality of care for patients with COPD.

Methods: This sequential explanatory mixed methods study was performed in two parts. In the quantitative part, a survey of 199 physiotherapists was conducted to evaluate their attitudes and knowledge, as well as the influence of contextual factors (i.e., practice policy and support from colleagues), in the implementation of the standard measurement set. In the qualitative part, 11 physiotherapists participated in individual interviews to elucidate their experiences using a thematical framework.

Results: The survey showed that, on average, 68.4% of the physiotherapists reported having a positive attitude about using the standard set, 85.0% felt they had sufficient knowledge of the measurement instruments, and 84.7% felt supported by practice policy and colleagues. In total, 80.3% of physiotherapists thought the standard set had added value in clinical practice, and 90.3% indicated that the measurement instruments can be valuable for evaluating treatment outcomes. The physiotherapists mentioned several barriers, such as lack of time and the unavailability of the entire standard set of measurement instruments in their practice. Moreover, the physiotherapists indicated that the measurement instruments have added value in providing transparency to policymakers through the anonymized publication of outcomes.

Conclusion: Physiotherapists support the use of the standard set of measurement instruments to improve the quality of physiotherapy treatment for patients with COPD.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a serious public health problem. This progressive disease affects the lungs, causing dyspnoea with exertion in particular, which has a negative effect on quality of life (1). Physiotherapy improves the quality of life of patients with COPD by increasing the physical capacity and decreasing breathlessness (2); thus, high-quality physiotherapy care for these patients is of high importance for achieving optimal treatment results. In recent years, routinely collected real-world data from electronic health records have become available from national data registries in the Netherlands. These data offer the opportunity to use patient outcomes in the interaction between the physiotherapist and the patient (e.g., in goal setting and shared decision-making), and to improve the quality of care by learning from aggregated outcomes within and between practices (3-5). Furthermore, routinely collected data may be used for external transparency, such as public reporting or pay-for-performance initiatives (6). It is important that valid outcomes and measurement instruments are selected, tested for their use in quality improvement, and validated by end users (7, 8).

Previous research investigated the barriers to and facilitators of physiotherapists using measurement instruments, revealing that they were not being routinely used (9-12). It was found that, although physiotherapists had a positive attitude towards the use of measurement instruments, they were not always sure which should be used for which patient. They indicated that a standard set of measurement instruments is needed, including instructions for their use and interpretation (7). The present lack of standardization in outcome measurements has meant that physiotherapy care approaches cannot be properly compared and evaluated (9).

A standard set of measurement instruments for Dutch physiotherapist practice, including patient-reported outcome measures (PROMs) and physical performance tests, was developed for patients with COPD and registered on the COMET website (13, 14). Physiotherapists can use this set for diagnostic purposes, goalsetting, and evaluating the outcomes of physiotherapy treatments for patients with COPD; however, it is unclear whether this standard set overcomes the described barriers for the successful implementation of routine data collection and the use of outcomes data to stimulate quality improvement.

Thus, the objectives of this study were 1) to evaluate the implementation of the set of measurement instruments for patients with COPD undergoing physiotherapy, and 2) to explore the perceptions of physiotherapists regarding the use of the set for goalsetting, quality improvement, and external transparency.

METHODS

Study design

A mixed methods approach with an explanatory sequential design was used by means of a survey and interviews with Dutch primary care physiotherapists. The standard set of measurement instruments was developed in a previous study (13), and included measures of the process and outcomes of physiotherapy care. Details of the set are available in Supplementary File 1. The set was implemented in a two-year time frame (January 2018 to December 2019) in 156 primary care practices, involving 295 physiotherapists (15). Twice a year, the participating practices received a report comparing their own collected data with benchmark data, presented in caterpillars plots (15).

The present study included two phases (see Figure 1). During the first phase, quantitative data from a survey of physiotherapists were analysed to evaluate their attitudes, knowledge, and the influence of contextual factors (i.e., practice policy and support from colleagues) in the use of the standard set for improving the physiotherapy treatments for patients with COPD. In the second phase, in-depth interviews were held with physiotherapists to gain a better understanding of their experiences of implementing the standard set of measurement instruments. The survey was executed from April–June 2018, while the interviews were conducted in March–June 2020.

The study protocol was approved by the Medical Ethical Committee of Radboud university medical centre (registration #2019-5455). The informed consent of each participant was obtained.

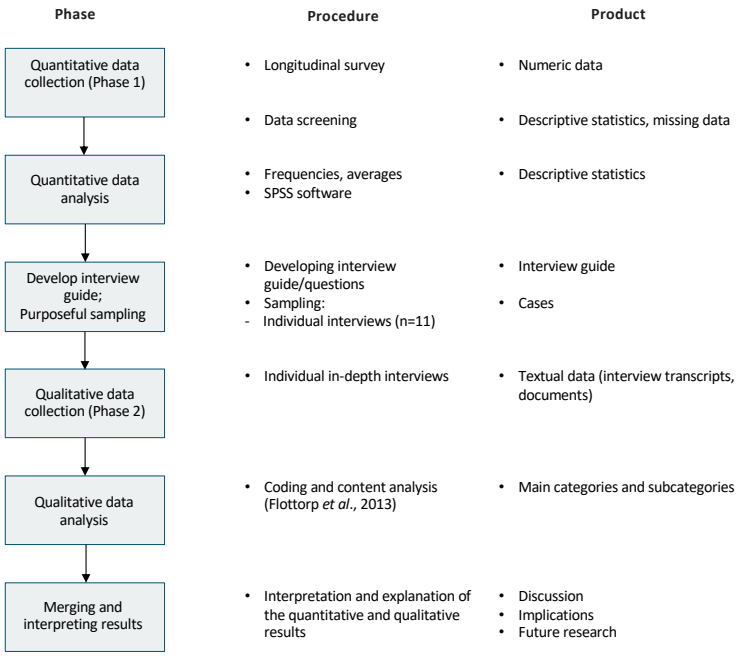


Figure 1 Diagram of the study. A sequential explanatory mixed method design.

Phase 1: Survey

Data collection

All physiotherapists who participated in the previous implementation study were invited to participate. A survey was sent via LimeSurvey version 2.06, with a total of three reminders. The survey was based on the previously developed 'PROM use self-assessment questionnaire' (3). The questions were allocated into three domains: attitude, knowledge, and context (3). The questionnaire asked for the (demographic) characteristics of each participant, their perceptions of the implementation of the standard set, their personal experience with the standard set, and the policy of their practice regarding the standard set. All questions were scored on a five-point Likert scale (1 = totally disagree, 5 = totally agree). Some minor changes were made to the original questionnaire because the current study used a specific standard set of measurement instruments instead of the general PROMs explored in the original article; for example, 'I know where to find PROMs' was changed to 'I know where to find the measurement instruments.'

Data analysis

Mean scores and standard deviations (SD) were calculated for each question and for the three domains. Descriptive statistics were calculated to describe the percentage

of physiotherapists who agreed on the questions provided. A score of 4 (agree) or 5 (totally agree) on a question was interpreted as agreed .SPSS version 25 was used for all calculations.

Phase 2: Interview study

The protocol for the interviews was based on the results of the survey. The questions were designed to identify the factors that potentially facilitate or hinder the use of the measurement instruments. Physiotherapists who had completed the survey in the first part of the study were eligible to participate in the interviews once their consent had been obtained. The physiotherapists were purposefully selected based on their demographic characteristics collected in the survey, including age, gender, working hours, and geographic location across the Netherlands. Data saturation was expected to be reached after 10–15 interviews. The COREQ checklist (16) was used as guidance for conducting all aspects of the qualitative research.

Data collection

Semi-structured in-depth interviews were pilot tested and conducted via video connection and audio recorded by one researcher (JZ). The interviews began with some general open-ended questions, after which the physiotherapists were asked about their experiences with the use of measurement instruments. Finally, the physiotherapists were asked about their perspectives on the potential use of the data for quality improvement and transparency (see Supplementary File 2 for the interview guide).

Data analysis

The interviews were transcribed verbatim. The transcripts were entered into Atlas.ti, a program used for analysing qualitative data, assigning codes, and allocating the codes into categories. We used a directed approach to content analysis [17]. We inductively coded the transcripts and then used an existing theoretical framework to guide higher-order clustering. Specifically, two researchers (JZ and AV) independently coded two transcripts and discussed the codes to reach a consensus. The remaining transcripts were coded by one researcher (JZ) and checked by another researcher (AV). During this process, new codes were added when needed after discussions between both researchers (JZ and AV). Based on their similarities, the codes were allocated into categories (by JZ and AV). Then we used the theoretical framework of Flottorp et al 2013[18] to cluster the codes to major categories and the seven domains of Flottorp: Guideline factors; Individual health professional factors; Patient factors; Professional interactions; Incentives and resources; Capacity for organizational change; and Social, political and legal factors. The theoretical framework of Flottorp facilitates the evaluation and reporting of tailored interventions. The clustering of the categories to the domains was discussed during meetings with all research members (JZ, AV, SvD, and PvdW) to reach a consensus. The research team (JZ, AV, SvD, and PvdW) also held several meetings throughout data collection

to discuss and interpret the preliminary findings, to make potential amendments to the interview guide, and to identify whether data saturation had been reached.

Trustworthiness

Both the quantitative and qualitative parts of this study are related to the validity of a mixed method design. A large sample size was used for the survey to minimize bias and possible validity threats. Interviews were held until data saturation was reached. The participants had no personal relationship with the researchers. The interview data were analysed by both JZ (a master's student) and AV (a physiotherapist and PhD student) to strengthen trustworthiness. On several occasions during the study, all research members (JZ, AV, PvdW (a physiotherapist and professor of allied health sciences), and SvD (a physiotherapist and senior researcher)) discussed the codes, categories, and domains to reach consensus about the findings from the interviews. The research members were trained in (and most had experience in) conducting qualitative research (AV, SvD, HK (a physiotherapist and senior researcher), and PvdW). Reliability and validity were established using the four components outlined by Guba and Lincoln (1981): credibility, transferability, dependability, and confirmability (19). The identities of the physiotherapists were considered confidential; therefore, the answers given by the physiotherapists during the interviews and in the survey were processed anonymously. Meaning that the transcripts of interviews in the current study cannot be linked to identities of participants by removing all identifiable information of the participants.

RESULTS

Survey

Of the 295 physiotherapists who participated in the implementation study, a total of 199 completed the survey (response rate: 67.4%). The mean age was 42.1 years (SD 12.0), and 92 participants were male (46.2%). The participating physiotherapists comprised a representative sample in terms of age and gender when compared with the national reference data (20). The mean number of hours worked per week among the male participants was 37.3 hours (SD 6.7 hours), whereas the mean working hours per week among the female participants was 28.4 hours (SD 6.1 hours). See Table 1 for full details.

Table 1 Characteristics of the physiotherapists participating in the survey.

	N	Age in years (SD)	Working hours per week (SD)
All participants	199	42.1 (12.0)	32.5 (7.7)
Male (%)	92 (46%)	43.6 (13.0)	37.3 (6.7)
Female (%)	107 (54%)	40.8 (10.9)	28.4 (6.1)

SD: standard deviation.

The results of the survey showed that the majority of respondents had positive opinions of the use of the measurement instruments and the implementation of the standard set. Table 2 provides a complete overview of the results of the survey per item and per category. Some items might have a slightly different response rate as not all participants answered all questions. Table 2 shows that 68.4% of the physiotherapists (in total) agreed with items related to having a positive attitude (mean score 3.88), 85% (in total) agreed with the items related to having sufficient knowledge (mean score 4.06), and a total of 84.7% agreed with the items related to context (mean score 4.16). This indicates that the highest gains in the implementation of the set of measurement instruments in clinical practice could be made by changing the attitude of the physiotherapists regarding the use of the standard set in daily practice. Of the physiotherapists who completed the survey, 91.7% agreed that the measurement instruments are useful in the evaluation of the treatment; 23.8% agreed that they would like to use the measurement instruments more often in clinical practice.

Table 2 Results of the survey on attitude, knowledge, and context.

	% (in total) who agree‡	**Mean (SD)**	**Min–max**§
Attitude	**68.4**	**3.88 (0.86)**	**1–5**
Using the measurement instruments helps me formulate a physiotherapeutic diagnosis	69.4	3.74 (0.86)	1–5
The measurement instruments are useful in the evaluation of the treatment	91.7	4.18 (0.69)	1–5
The measurement instruments have a positive influence on the quality of physiotherapy healthcare	74.1	3.83 (0.83)	1–5
It is important to register patient experiences objectively with the measurement instruments	87.6	4.07 (0.96)	1–5
Using the measurement instruments in clinical practice does takes too much time†	51.7	4.60 (1.00)	1–5
I would like to use the measurement instruments more often in clinical practice	23.8	2.82 (0.96)	1–5
I have experienced the added value of the measurement instruments in clinical practice	80.8	3.96 (0.72)	1–5
Knowledge	**85.0**	**4.06 (0.72)**	**1–5**
I know where to find the measurement instruments	93.3	4.31 (0.76)	1–5
I am capable of using the measurement instruments with my patients	93.7	4.31 (0.71)	1–5
I am able to interpret the results of the measurement instruments	91.7	4.19 (0.69)	1–5
Using the measurement instruments does not affect my professional authority to make my own decisions	79.3	3.84 (0.79)	1–5
All patient needs can be registered in the measurement instruments	50.2	3.40 (0.77)	1–5

Table 2 Continued

	% (in total) who agree‡	Mean (SD)	Min–max§
I am able to use the measurement instruments within physiotherapeutic methodical action	93.8	4.17 (0.62)	1–5
I use the measurement instruments in daily practice	93.3	4.26 (0.68)	1–5
Context	**84.7**	**4.16 (0.74)**	**1–5**
The use of the set measurement instruments fits with how I am used to working	71.0	3.70 (0.74)	2–5
The measurement instruments are available in my practice	92.4	4.41 (0.67)	1–5
In our practice, we have made arrangements for how to use the measurement instruments	84.4	4.09 (0.85)	1–5
My supervisor(s) supports the employees in the use of measurement instruments	84.3	4.41 (0.78)	1–5
My supervisor(s) use the measurement instruments in clinical practice themselves	86.0	4.10 (0.78)	1–5
My supervisor(s) requires employees to report digitally using the measurement instruments	88.6	4.20 (0.79)	1–5
My colleagues also use the measurement instruments in clinical practice	83.2	4.25 (0.63)	1–5
In our practice, the use of measurement instruments fits well in the way of working	88.0	4.15 (0.73)	1–5

† Since all items should have the same scoring procedure, this item was recoded positively.
‡ Score of 5 (totally agree) or 4 (agree).
§Scored on a five-point Likert scale 1 = totally disagree, 5 = totally agree

Interviews

In total, 11 interviews were held. After discussing the preliminary results of 10 interviews, the researchers concluded that one more interview was needed to be certain that data saturation was reached. The interviews took between 30 and 70 minutes. Six of the interview participants were male (54.5%) with a mean age of 39.5 years, while the females (45.5%) had a mean age of 37.6 years. An overview of the characteristics of the participants is outlined in Supplementary File 3.

After analysing the data from the interviews, the codes were clustered into eight major categories: 1) Applicability and time frame of assessments of the measurement instruments in the standard set; 2) Knowledge and skills of physiotherapists; 3) Acceptance (including attitudes) of physiotherapists; 4) Patient motivation and behaviour; 5) Quality improvement; 6) Information system of the practice; 7) Availability of resources in the practice; 8) Transparency. These major categories were allocated to the seven generic domains identified by Flottorp *et al.* (21) (see Table 3). The categories are described in detail in the following paragraph.

Table 3 Categorization of the generic and specific domains, major categories, and codes.

Generic domains according to Flottorp *et al.* (21)	Major categories	Codes
Guideline factors	1) Applicability and time frame of assessments of the measurement instruments in the standard set	Goals of using the measurement instruments
		Barriers to using the measurement instruments
		Facilitators of using the measurement instruments
		Presented information for using the standard set is sufficient
Individual health professional factors	2) Knowledge and skills of physiotherapists	Different experiences of using the measurement instruments related to additional COPD training
	3) Acceptance (including attitudes) of physiotherapists	Different experiences of using the measurement instruments related to age
		Mixed perspectives on different ways to use the data of the measurement instruments
Patient factors	4) Patient motivation and behaviour	Resistance to frequent measuring
		Interest in own results
		Participating in filling in questionnaires
		Enthusiasm towards using an accelerometer
		Barriers to using the accelerometer
Professional interactions	5) Quality improvement	Barriers to quality improvement
		Facilitators of quality improvement
		Barriers of having a small practice and little capacity
		Feedback on the measurement instruments results is valued
Incentives and resources	6) Information system of the practice	Barriers to the implementation
		Facilitators of the implementation
		Software problems
		Software facilitators
	7) Availability of resources in the practice	Lack of space to complete the 6MWT
		Microfet™ is expensive to purchase
		Shortage of accelerometers
Capacity for organizational change	6) Information system of the practice	Barriers to the implementation
		Facilitators of the implementation
		Software problems
		Software facilitators
	7) Availability of resources in the practice	Lack of space to complete the 6MWT
		Microfet™ is expensive to purchase
		Shortage of accelerometers

Table 3 Continued

Generic domains according to Flottorp *et al.* (21)	Major categories	Topics
Social, political, and legal factors	8) Transparency	Positive perspectives towards making the anonymized standard set data transparent for policymakers
		Negative perspectives towards making the standard set data transparent at an individual level

6MWT: six-minute walk test; COPD: chronic obstructive pulmonary disease.

1) Applicability and time frame of assessments of the measurement instruments in the standard set

Generally, the physiotherapists stated that the use of the measurement instruments in the standard set was feasible because they are sufficient to provide insight into the effect of the treatment without taking too much time to complete. The physiotherapists also indicated that the standard set was able to measure what is necessary to be able to evaluate and reorganize future treatment sessions based on the outcome, which is one of the goals of the use of the standard set:

> *"It [the standard set] guides your therapy and treatment plan and that of course has the effect that you have a better treatment plan for the patient and hopefully a better result" [I.09].*

Despite the generally positive experience with the standard set, barriers were also identified for specific measurements. According to some physiotherapists, the Microfet™ was unreliable because it depends on the way in which it is used, as well as being affected by the experience of the physiotherapist:

> *"With the Microfet™, there is a difference in testing. There is too much of a difference in the outcomes between individuals [physiotherapists]. It is just very 'sensitive' to the way in which it is used" [I.10].*

In general, the participants agreed that an accelerometer provides valuable insights into the general activity of the patient; however, some physiotherapists stated that the accelerometer is not accurate in estimating the number of steps per day:

> *"I always have doubts about the accuracy of the accelerometer, but it does give an indication […] Some patients still score very few steps, which gives me a good insight that I should speak to them to see how I can encourage them to move more. Otherwise, you would have no insight in that area" [I.06].*

It was indicated in the instructions for the standard set that measurements should be conducted every three months; however, the physiotherapists stated that this was not always possible due to a lack of time or the status of the patient.

> *"[…] and sometimes patients have a bad day the day you were planning to measure the performance measures from the standard set, at those moments they are absolutely not motivated. Then it is difficult for me to tell them that we still need to perform the measures" [I.9]*

During the interviews, variation was observed in the frequency at which the physiotherapists used the standard set; while some physiotherapists repeated the measurements on schedule, other physiotherapists reported using the measurement instruments every six months. In general, however, the standard set was found to be very useful. The participants commented that the standard set ensures that they measure consciously:

> *"Yes, I still use the standard set. With the standard set I learned to structurally measure outcomes. When it is really busy at work, and you think that you do not have time, then the standard set motivates me to measure the repeated measurement." [I.10]*

Generally, the physiotherapists thought that the information provided before and during the project was useful and easy to apply. The participants commented that the protocol was simple to implement and follow.

Barriers associated with the instructions were also mentioned during the interviews, however. Some physiotherapists indicated that when specific measurement instruments are not available (e.g., the Microfet™) or cannot be performed exactly according to the instructions (e.g., no ten-metre space available for the 6MWT), alternative measurement instruments or instructions should be given. Moreover, according to the physiotherapists, the fact that some measurement instruments are optional should be made clearer in the information provided for the standard set:

> *"I understood that [some instruments are optional], but colleagues of mine asked: 'we should also take that test, right?' 'Well, that is not necessary with this client because that is not a goal. His strength is already good, so you don't need to test that further'. It was in the text [information for the standard set], but maybe mention it more often in several places or something" [I.01].*

2) Knowledge and skills of physiotherapists

The interviews revealed that physiotherapists who had not received additional COPD-specific training had less of a positive experience with the use of the standard

set of measurement instruments because they lacked the underlying knowledge of these procedures. According to the participants, however, skills are related to the experience of the physiotherapists. The participants also stated that specialized physiotherapists should continuously develop their knowledge by undertaking additional training to keep themselves more alert about their clinical process:

> *"I always find that when I have completed a COPD training course, I am more up-to-date and alert. But I think that applies to everyone" [I.07].*

3) Acceptance (including attitudes) of physiotherapists

The participants indicated that the younger generation of physiotherapists are trained in using measurement instruments and reporting the data, and therefore have more positive experiences of using them than the older generation. The critical attitudes of older physiotherapists were noted as a barrier for using measurement instruments, as this group is less familiar with them. The physiotherapists indicated that the use of measurement instruments comes with too much reporting, which is time consuming:

> *"I know a lot of colleagues in my age group who think that it is all nonsense [the use of measurement instruments] and do not want to explore the use of measurement instruments and start working with them. And yes, that is a pity" [I.04].*

4) Patient motivation and behaviour

The physiotherapists mentioned that the motivation of the patient is built on providing sufficient information about the importance of using measurement instruments. The participants indicated that some patients were not motivated to complete the measurements every three months because they are not used to routine testing; however, most of the patients were interested in their results and were therefore more motivated to complete the standard set of measurements:

> *"What is striking is that the patients also like to evaluate the results every three to four months, to do all the tests and measurements. They are also interested to see how they are doing, not only in the function of their lung" [I.06].*

The physiotherapists mentioned that the limited number of questionnaires included in the standard set meant the patients had no problems completing them. The ability to complete the questionnaires online also motivated the patients because it takes less time. The use of instruments that allow patients to track their activity made them more keen to complete the measurements and tests. According to the participants, this was because a goal (amount of steps) was given to the patients:

> *"I do see, when the patients get such a goal, that they like it. They say things like: 'oh, I've taken 5000 steps, let's try to set 5500 or so this week'" [I.02].*

5) Quality improvement

Both barriers and facilitators were mentioned for using aggregated outcomes for quality improvement purposes by comparing outcomes between physiotherapists. Although this goal is valued, the physiotherapists experience the use of the data as confrontational because the scores are compared between physiotherapists, and some have higher scores than others:

> *"It is very confrontational for the treating physiotherapists and they defend themselves. I myself also tend to do it, because you sometimes feel more or less attacked. It shouldn't be like that; it has to be for learning, it has to be for improvement" [I.07].*

The physiotherapists indicated that the results of the measurement instruments should be case-mix corrected for the burden of COPD, because this condition explains the results to a large extent. This will facilitate the use of the data for quality improvement:

> *"If, for example, one physiotherapist treats more patients in classes A and B [burden of disease] and the other more from C and D, what do these data say then?" [I.12].*

The physiotherapists stated that the use of data for quality improvement initiatives also depends on the size and capacity of the practice. Most small practices employ less specialized physiotherapists, who therefore receive less feedback from other physiotherapists with the same specialism and have fewer colleagues with whom to compare their data. The participants who were the only COPD-specialized physiotherapist of their practice indicated that they were curious about their own outcomes and willing to compare outcomes with colleagues:

> *"We can use the graphs to see whether there is a difference [between the scores of physiotherapists], and then we can explore where that difference comes from. I believe in that way we can learn from each other" [I.09].*

To facilitate the use of the data from the measurement instruments for quality improvement, it is useful for the physiotherapists to be able to record specific factors, such as changes in medication or hospitalization. In that way, the cause of the possible variation between physiotherapists within a practice could be identified more easily.

The visual feedback of the results of the standard set in caterpillar plots was highly valued by the participants, who felt they could easily use the data to compare their results with those of other physiotherapists.

Moreover, the physiotherapists enthusiastically indicated that they want to use the data in their practice to improve their quality. As the data would be anonymously provided, they indicated that practices must be able to contact other practices with better scores to be able to learn how to improve their own quality without violating the privacy of other practices.

A major barrier for using the feedback of the results is that some physiotherapists received feedback with less data than they had sent. The physiotherapists indicated that they would therefore have appreciated receiving more feedback then twice a year a feedback report comparing their own collected data with benchmark data report. For example, the appreciated to be informed when their data has been received:

> *"These data would be nice to present between the feedback moments. You [the person who receives the data] could ask after six months or a year: 'we now have received this number of measurements. Is this in line with the number of patients you treat and for whom you have taken the measurements?' If it is not correct, then you can try to find the reason behind it" [I.06].*

6) Information system of the practice

Most physiotherapists indicated that they did not experience problems with the way the standard set was implemented, nor with the software they were using:

> *"A protocol for the standard set was just assigned, right? So, we could actually just implement that" [I.09].*

Other physiotherapists mentioned several missed opportunities concerning the implementation of the protocol of the standard set and the software; for example, some participants indicated that their practice found it difficult to correctly implement the standard set at the beginning of the study. Most of the physiotherapists mentioned that this was due to the way their practice leader or colleague had informed them about how to find the standard set in their electronic health record:

> *"We were not informed correctly, as that colleague [who informed the others about the standard set] actually started that trajectory before fully implementing the standard set in our system. They just told us what the intention was and how we should start with it" [I.05].*

7) Availability of resources in the practice

Most participants commented that all measurement instruments, and the resources required to properly use them, were available; however, a few participants stated that some measurement instruments were not available in their practice. This was

mainly true for the Microfet™, a tool to measure muscle strength, as this instrument is expensive to purchase:

> *"So the hand-held dynamometer, we don't have that in our practice. As an investment it is quite expensive. So, eventually we never decided to buy the Microfet™" [I.08]*

Also, some of the participants mentioned that some practices do not have enough space to optimally use some of the measurement instruments. This specifically holds true for the six-minute walk test (6MWT), for which it is advised that the patients walk ten metres back and forth in a straight line, but this is not possible in every practice.

8) Transparency

Another aim of the project was to make the anonymous results of the measurement instruments transparent for stakeholders and eventually to make the data totally transparent on physiotherapist or practice level. All physiotherapists indicated that the transparency of the data is an important factor for improving the quality of care; however, the participants indicated that the data collection should be optimized before it is used for external transparency purposes. All physiotherapists stated that it is important to perform a case-mix correction for the burden of COPD. Furthermore, the physiotherapists indicated that it is important to harmonize the use of the measurement instruments:

> *"I think it is good to compare between different practices, but it is not enough with the measurements we use now because, for example, the six-minute walk test can be measured in many different ways" [I.09].*

Most of the physiotherapists indicated having no problem with the data being accessible in an anonymous form for policymakers when it is case-mix corrected, as mentioned above:

> *"I think that the more information we can provide to policymakers, the better the directives they write" [I.06].*

The physiotherapists mentioned that making the data of the measurement instruments totally transparent on an individual level is important, yet they expressed some concerns related to potential gaming because they fear the negative (financial) consequences. A third party would therefore be needed to perform the measurements, according to some participants. Another barrier mentioned was that the physiotherapists think that both patients and health insurers might misinterpret the results:

"It is always difficult to know how another party would interpret such data. You may want to be transparent because it is important to you, but I am not sure that the patient who reads it can interpret it correctly" [I.10].

DISCUSSION

The results of our study show that the participating physiotherapists in the survey had a positive attitude towards, felt knowledgeable about, and were supported by practice policy and colleagues in the use of a standard set of measurement instruments with patients with COPD. The qualitative analysis resulted in experiences of physiotherapists with implementing the standard set of measurement instruments into eight defined categories. Although some barriers were mentioned, the physiotherapists during the interviews valued using the measurement instruments on patient-level for the evaluation of physiotherapy treatments, as well as on aggregated-level for quality improvement purposes. Moreover, the physiotherapists indicated that the measurement instruments have added value for the anonymized publication of outcomes, providing transparency to policymakers. To our knowledge, this was the first study that used a mixed methods design to evaluate the experiences of physiotherapists regarding the implementation of a standard set of measurement instruments for the improvement of primary care physiotherapy treatments for patients with COPD.

In accordance with the present evaluation of the implementation of a standard set, previous studies have identified barriers for implementing a guideline for COPD physiotherapy treatment (22). Similarities were found in both the positive attitude of physiotherapists towards using measurement instruments and the negative finding that using measurement instruments takes too much time (22). In more general studies of the use of measurement instruments in physiotherapy, lack of time was again reported as a barrier (9, 23-25).

Another important finding was that patients are more motivated to undertake the tests when sufficient information about the importance of the measurement instruments is provided by the physiotherapists. This is consistent with the study of Østergaard *et al.* (26), in which the patients were found to be less active when physiotherapists did not provide information about the importance of physical activity.

To identify the experiences of primary care physiotherapists regarding the use of measurement instruments, most researchers only used surveys and focussed on specific outcome instruments (23, 27). The present study used a survey combined with semi-structured interviews to provide additional information and explanations to the answers given in the survey. Moreover, this design was used to create a complete

overview of the experience of using all the measurement instruments, including the use of the data for quality improvement and transparency.

Prior studies have also developed standard or core sets of outcome measures for patients with COPD (28-32). Most of these sets were to be used in clinical trials (28, 29, 31) or were not designed for the evaluation of the physiotherapy treatment of patients with COPD (30, 32). In the current study, we evaluated the implementation of the standard set that was developed for use in Dutch primary physiotherapy care. We believe that researchers, policymakers, and other stakeholders can learn from the experiences of physiotherapists using the standard set of measurements to collect aggregated outcomes for quality improvement and external transparency.

Limitations

This study has several limitations. First, interviews were held by only one of the researchers (JZ) and member checking was not performed. This could have negatively influenced the trustworthiness and validity of this study (33). To strengthen trustworthiness, the interviews were independently analysed, and codes were assigned by JZ and AV.

Second, as only physiotherapists were interviewed in this study, it is important to indicate that the barriers and facilitators allocated to the domain 'patient factors' were based on the perception of the physiotherapists and not obtained from the patients themselves.

Lastly, the major categories that emerged in this study were allocated to the seven domains developed by Flottorp *et al.* (21); however, the domains 'incentives and resources' and 'capacity for organizational change' both contained the same two major categories because no distinction could be made when allocating the different topics and categories to the domains. Despite this issue, data saturation was obtained and a consensus about the findings was reached as all research members discussed the codes, categories, and domains several times during the study.

Implications for practice

Feedback regarding the outcome data might promote quality improvement, but its effectiveness is related to how the feedback is provided (34). The physiotherapists indicated that the feedback is very useful for quality improvement; thus, it can be assumed that feedback regarding the measurement instruments can contribute to quality improvement initiatives. This could lead to better physiotherapy treatment for patients with COPD; however, future research should explore how the use of measurement instruments affects the quality of physiotherapy treatment and the outcomes of care.

The physiotherapists indicated that they were sceptical and not prepared to provide their data for full transparency yet, because they believed it likely that other physiotherapists would manipulate their outcomes to avoid negatively affecting their reimbursement by health insurers. The participants therefore suggested that the measurements should be performed by a third party. This idea should be explored in the future before making the outcomes totally transparent.

Sets of measurement instruments are always subject to change, and the routine evaluation of the instruments is always necessary. We will therefore routinely discuss, improve, implement, and evaluate the standard set in future research.

Conclusion

This mixed method study shows that the participating physiotherapists supported the use of a standard set of measurement instruments to improve the quality of physiotherapy treatment for patients with COPD. Eight categories were identified in the physiotherapist experiences with the use of the standard set for these patients. Moreover, we showed that the routine use of the set of measurement instruments has the potential to be used for the anonymized publication of outcome data, providing transparency to policymakers. The results of this study could be used for future projects focussed on improving, implementing, and evaluating the standard set.

Acknowledgements

The physiotherapists who contributed to the interviews are acknowledged for their effort and for sharing their experiences.

REFERENCES

1. Hu J, Meek P. Health-related quality of life in individuals with chronic obstructive pulmonary disease. Heart Lung. 2005;34(6):415-22.
2. Dimitrova A, Izov N, Maznev I, Vasileva D, Nikolova M. Physiotherapy in Patients with Chronic Obstructive Pulmonary Disease. Open Access Maced J Med Sci. 2017;5(6):720-3.
3. Meerhoff GA, van Dulmen SA, Maas MJM, Heijblom K, Nijhuis-van der Sanden MWG, Van der Wees PJ. Development and Evaluation of an Implementation Strategy for Collecting Data in a National Registry and the Use of Patient-Reported Outcome Measures in Physical Therapist Practices: Quality Improvement Study. Phys Ther. 2017;97(8):837-51.
4. Porter ME. What is value in health care? The New England journal of medicine. 2010;363(26):2477-81.
5. Van Der Wees PJ, Nijhuis-Van Der Sanden MW, Ayanian JZ, Black N, Westert GP, Schneider EC. Integrating the use of patient-reported outcomes for both clinical practice and performance measurement: views of experts from 3 countries. Milbank Q. 2014;92(4):754-75.
6. Van Der Wees PJ. Framework for the use of PROMs. In: Recommendations to OECD Ministers of Health from the high level reflection group on the future of health statistics. Paris: OECD. 2017;27.
7. van der Wees PJ, Verkerk EW, Verbiest MEA, Zuidgeest M, Bakker C, Braspenning J, et al. Development of a framework with tools to support the selection and implementation of patient-reported outcome measures. J Patient Rep Outcomes. 2019;3(1):75.
8. National Quality Forum I, editor Patient reported outcomes (PROs) in performance measurement. Washington (DC): National Quality Forum; 2013.
9. Swinkels RA, van Peppen RP, Wittink H, Custers JW, Beurskens AJ. Current use and barriers and facilitators for implementation of standardised measures in physical therapy in the Netherlands. BMC Musculoskelet Disord. 2011;12:106.
10. Van Peppen RP, Maissan FJ, Van Genderen FR, Van Dolder R, Van Meeteren NL. Outcome measures in physiotherapy management of patients with stroke: a survey into self-reported use, and barriers to and facilitators for use. Physiother Res Int. 2008;13(4):255-70.
11. Kyte DG, Calvert M, van der Wees PJ, ten Hove R, Tolan S, Hill JC. An introduction to patient-reported outcome measures (PROMs) in physiotherapy. Physiotherapy. 2015;101(2):119-25.
12. Eldin Alreni AS, McRobert C, McLean SM. Utilisation of outcome measures in the management of non-specific neck pain: A national survey of current physiotherapy practice in the UK. Musculoskelet Sci Pract. 2021;52:102347.
13. Verburg AC, van Dulmen SA, Kiers H, Ypinga JH, Nijhuis-van der Sanden MW, van der Wees PJ. Development of a Standard Set of Outcome Domains and Proposed Measures for Chronic Obstructive Pulmonary Disease in Primary Care Physical Therapy Practice in the Netherlands: a Modified RAND/UCLA Appropriateness Method. International journal of chronic obstructive pulmonary disease. 2019;14:2649-61.
14. Heeren J, Triemstra M, Boer D. Ontwikkeling van de PREM Fysiotherapie; een vragenlijst over de ervaren kwaliteit van de fysiotherapeutische zorg 2016 [Available from: https://www.nivel.nl/sites/default/files/bestanden/Ontwikkeling-PREM-Fysiotherapie.pdf.

15. Verburg ACvD, S. A.; Kiers, H.; Nijhuis-van der Sanden, M. W. G.; van der Wees, P. J. A practice test and selection of a core set of outcome-based quality indicators in Dutch primary care physical therapy for patients with COPD: a cohort study. Submitted.

16. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care. 2007;19(6):349-57.

17. Hsieh H-F, Shannon SE. Three Approaches to Qualitative Content Analysis. Qualitative Health Research. 2005;15(9):1277-88.

18. Flottorp SA, Oxman AD, Krause J, Musila NR, Wensing M, Godycki-Cwirko M, et al. A checklist for identifying determinants of practice: A systematic review and synthesis of frameworks and taxonomies of factors that prevent or enable improvements in healthcare professional practice. Implementation Science. 2013;8(1):35.

19. Guba EG, Lincoln YS. Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches: Jossey-Bass; 1981.

20. Dool Jvd, Schermer, T. Zorg door de fysiotherapeut. Jaarcijfers 2018 en trendcijfers 2014-2018. Utrecht: Nivel. 2019.

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

22. van der Wees PJ, Zagers CA, de Die SE, Hendriks EJ, Nijhuis-van der Sanden MW, de Bie RA. Developing a questionnaire to identify perceived barriers for implementing the Dutch physical therapy COPD clinical practice guideline. BMC Health Serv Res. 2013;13:159.

23. Jette DU, Halbert J, Iverson C, Miceli E, Shah P. Use of standardized outcome measures in physical therapist practice: perceptions and applications. Phys Ther. 2009;89(2):125-35.

24. Stokes EK, O'Neill D. Use of outcome measures in physiotherapy practice in ireland from 1998 to 2003 and comparison to canadian trends. Physiother Can. 2008;60(2):109-16.

25. van Tuijl AAC, Wollersheim HC, Fluit C, van Gurp PJ, Calsbeek H. Development of a tool for identifying and addressing prioritised determinants of quality improvement initiatives led by healthcare professionals: a mixed-methods study. Implement Sci Commun. 2020;1:92.

26. Ostergaard EB, Sritharan SS, Kristiansen AD, Thomsen PM, Lokke A. Barriers and motivational factors towards physical activity in daily life living with COPD - an interview based pilot study. Eur Clin Respir J. 2018;5(1):1484654.

27. Haigh R, Tennant A, Biering-Sorensen F, Grimby G, Marincek C, Phillips S, et al. The use of outcome measures in physical medicine and rehabilitation within Europe. J Rehabil Med. 2001;33(6):273-8.

28. Camus-García E, González-González AI, Heijmans M, Niño de Guzmán E, Valli C, Beltran J, et al. Self-management interventions for adults living with Chronic Obstructive Pulmonary Disease (COPD): The development of a Core Outcome Set for COMPAR-EU project. PLOS ONE. 2021;16(3):e0247522.

29. Cazzola M, MacNee W, Martinez FJ, Rabe KF, Franciosi LG, Barnes PJ, et al. Outcomes for COPD pharmacological trials: from lung function to biomarkers. European Respiratory Journal. 2008;31(2):416-69.

30. Cieza A, Ewert T, Ustün TB, Chatterji S, Kostanjsek N, Stucki G. Development of ICF Core Sets for patients with chronic conditions. Journal of rehabilitation medicine. 2004(44 Suppl):9-11.

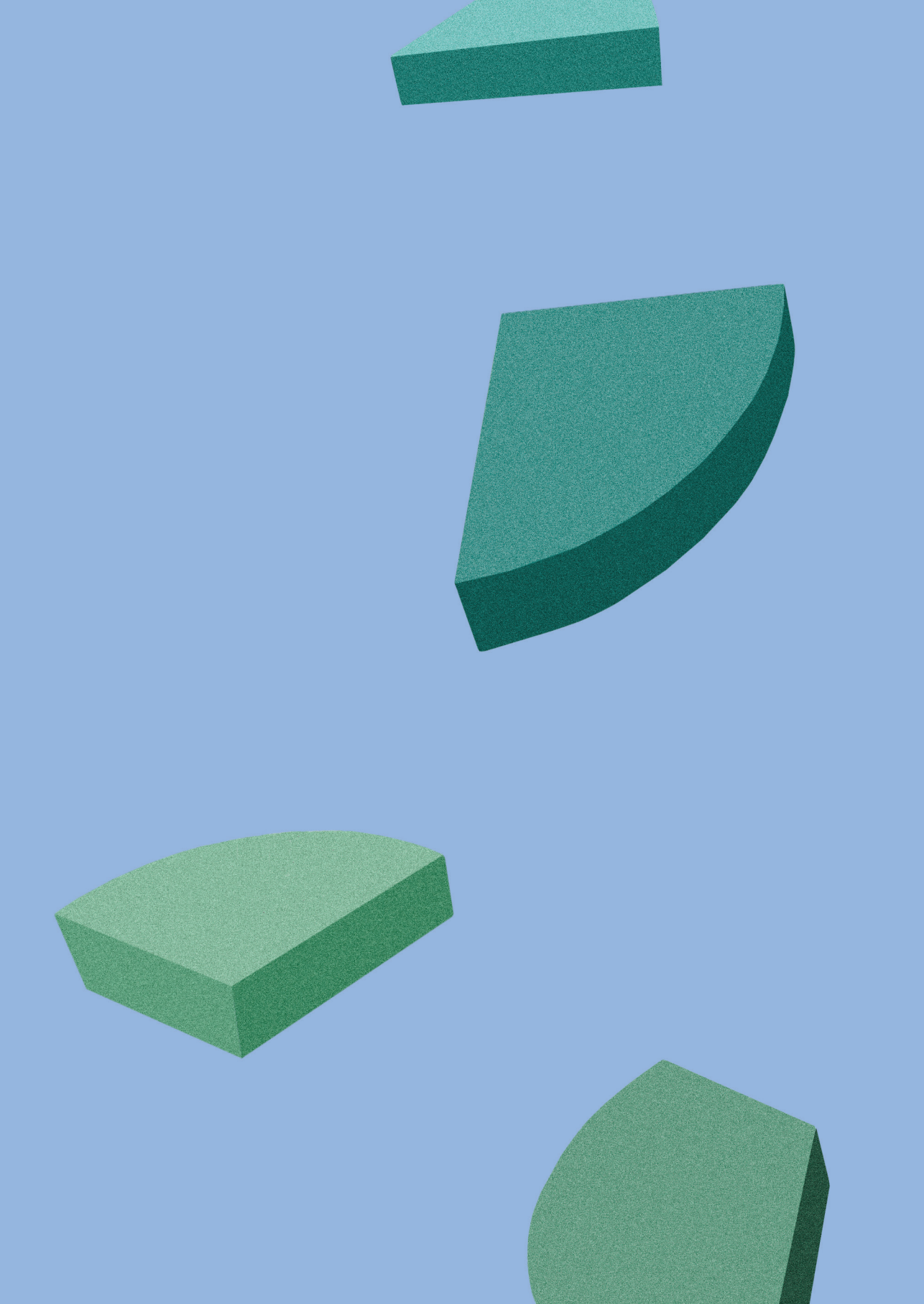
31. Mathioudakis AG, Abroug F, Agusti A, Bakke P, Bartziokas K, Beghe B, et al. Core outcome set for the management of acute exacerbations of chronic obstructive pulmonary disease: the COS-AECOPD ERS Task Force study protocol. ERJ Open Research. 2020;6(3):00193-2020.

32. Souto-Miranda S, Marques A. Triangulated perspectives on outcomes of pulmonary rehabilitation in patients with COPD: a qualitative study to inform a core outcome set. Clinical rehabilitation. 2018:269215518821405.

33. Pitney WA. Strategies for establishing trustworthiness in qualitative research. International Journal of Athletic Therapy and Training. 2004;9(1):26-8.

34. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. The Cochrane database of systematic reviews. 2012(6):Cd000259.

35. Salome PL, in 't Veen JC, Euving E, Göttgens W, Harms LM, de Jong JB, et al. Dutch Lung Alliance (LAN) COPD Zorgstandaard 2016.



CHAPTER 7

General discussion

In this chapter, the main findings of this thesis will be presented and discussed in the context of the relevant literature. The overarching aim of this thesis was to develop, select and test a core set of outcome-based quality indicators for physical therapy practice for patients with non-specific low back pain (NSLBP) and chronic obstructive pulmonary disease (COPD), based on patient-reported outcome measures (PROMs) and physical performance measures. The goal was that the analysis of the combined results of the core set could be publicly reported and used to improve clinical care, such as was presented in figure 1 of chapter 1, and to compare and discriminate treatment outcomes in physical therapy practice, i.e., between physical therapists or physical therapy practices. First, we point out the main findings of this thesis, followed by a reflection on the five themes we identified: 1) data collection, 2) the development of outcome-based quality indicators, 3) using outcomes in clinical decision-making, 4) using quality indicators for internal quality improvement, and 5) using quality indicators for external transparency. By means of these themes, we will give an interpretation of the results of our studies, discuss methodological considerations and provide recommendations for practice and research. Finally, we will formulate our overall conclusions.

MAIN FINDINGS

In chapter 2, we described the development of a standard set of six clinical outcome domains and associated measures for use with patients with NSLBP in primary care physical therapy practice, including the selection of a tool to stratify patients into subgroups. In chapter 3, we described the development of a standard set of outcome domains and associated measures for patients with COPD, consisting of four mandatory measures, two conditional measures depending on the treatment goal and two exploratory measures. Furthermore, a measure was included to identify subgroups based on the burden of disease. The development processes of the standard sets for NSLBP and COPD were roughly the same. Both consensus-based standard sets were accepted to be relevant and feasible by stakeholders (i.e., patients and patient associations, physical therapists, researchers, policy-makers and health insurers) and were deemed useful for a) the interaction between the patient and healthcare professional, e.g., for shared decision-making in goal setting and for monitoring and feedback based on outcomes, b) internal quality improvement, and c) the external transparency of primary care physical therapy practices. Both sets provide a promising basis for the further development of quality indicators in physical therapy practice to improve clinical care and public reporting.

In chapter 4, we described the development, selection and testing of a core set of six outcome-based quality indicators for patients with NSLBP, based on the standard set from chapter 2, the cohort data and the consensus in focus groups of physical therapists and researchers. Our analysis show that the comparability of the

quality indicators increased after case-mix adjustment for age, gender, chronicity (duration of complaints) and a baseline score of the outcome measures. The discriminability of the outcome-based quality indicators between physical therapy practices was estimated to be adequate. We defined comparability as the extent to which the outcomes of the quality indicator are comparable between practices, and discriminability as the extent to which the outcomes of the quality indicator are able to discriminate between practices. The outcome-based quality indicators were accepted by stakeholders as having added value in daily practice, as well as for quality-improvement purposes.

In chapter 5, we used the standard set from chapter 3 as the basis for development, selection and testing of a core set of seven outcome-based quality indicators for patients with COPD, showing that the comparability of all outcome-based quality indicators for patients with COPD increased after case-mix adjustment for age, gender and the baseline scores of the measures. The discriminability of outcomes between physical therapy practices fluctuated; for six of the 11 indicators, the discriminability could not be interpreted as adequate. All participants of the focus groups reached consensus on the selection of the core set, and perceived that it added value for quality-improvement purposes.

The studies described in chapters 4 and 5 showed that not all physical therapists were able to collect enough data. We identified several reasons why the data collection could be insufficient, such as technical issues or the engagement of end-users (i.e., patients and physical therapists). Future efforts should highlight the usefulness of data collection for end-users and improve the validity of the data collected.

Finally, in chapter 6, we aimed to evaluate the experiences of physiotherapists with the implementation of a standard set of outcome measures for patients with COPD. We showed that, according to physical therapists, the measurement instruments (PROMs and physical performance measures) of the standard set have added value for the physical therapy treatment of patients with COPD. Although some barriers were mentioned, physical therapists valued using the measurement instruments in clinical decision-making for goal setting and for the evaluation of physical therapy treatments for patients with COPD. Measuring outcomes with the measurement instruments was also valued for quality-improvement purposes, both for enabling the interpretation of their own (aggregated) outcomes and comparing outcomes with peers. Moreover, the physical therapists indicated that the quality indicators may have added value for the (anonymised) publication of outcomes, providing transparency to policy-makers. The physical therapists indicated that they were sceptical and not prepared to provide their data for full external transparency yet because they were concerned about the consequences for reimbursement by health insurers.

During the project we formed a steering committee with representatives of associations for patients with COPD (Longfonds) and NSLBP (Dutch association for low back pain "The Spine"), policy-makers of professional physical therapy bodies (the Royal Dutch Society for Physical Therapy (KNGF) and the Association for Quality in Physical Therapy (SKF)) and health insurers (CZ and the Friesland). During the meetings, we discussed the views and perspectives of stakeholders regarding the value and implementation of outcome-based quality indicators for Dutch physical therapy. The committee monitored the process and finally accepted the core sets of outcome-based quality indicators for patients with COPD and NSLBP.

Table 1 Participants, goals and results of each chapter in this thesis

Chapter	Goal	Participants	Results
Chapter 2	To develop a standard set of measures for patients with NSLBP in primary care physical therapy practice	Patients with NSLBP and representatives from the patient association, physical therapists, researchers, policy-makers and health insurers	A standard set of six outcome domains and measures was accepted as being relevant and feasible by stakeholders, and deemed useful for a) interactions between patients and healthcare professionals, b) internal quality improvement and c) external transparency
Chapter 3	To develop a standard set of measures for patients with COPD in primary care physical therapy practice	Patients with COPD and representatives from the patient association, physical therapists, researchers, policy-makers and health insurers	A standard set of eight domains and associated measures was accepted to be relevant and feasible. The set can be useful for a) interactions between patients and healthcare professionals b) internal quality improvement and c) external transparency
Chapter 4	To develop, select and test a core set of outcome-based quality indicators, accepted by stakeholders for their usability and perceived added value as quality-improvement tools	Patients with NSLBP, physical therapists and senior researchers	After describing the comparability and discriminability of the indicators, stakeholders selected a final core set of six quality indicators for patients with NSLBP in primary care physical therapy practice
Chapter 5	To develop, select and test a core set of outcome-based quality indicators that are well-accepted by physical therapists based on their perceived added value as quality-improvement tools	Patients with COPD, physical therapists and senior researchers	The comparability and discriminability of the indicators was described. All participants in the focus groups accepted the quality indicators as a quality-improvement tool based on their perceived added value, and selected a core set of seven outcome-based quality indicators for patients with COPD

Table 1 Continued

Chapter	Goal	Participants	Results
Chapter 6	To explore the implementation of the core set developed in chapter 5 for quality improvement and public reporting, based on the experiences of physical therapists	Physical therapists who treated patients with COPD	Although some barriers were mentioned, physical therapists valued using the standard set for the evaluation of their treatments. The set was perceived to be useful for quality-improvement purposes, for the (anonymised) public reporting of outcomes, and for providing transparency to policy-makers

Data collection

Interpretation of the results

One aspect that can enhance the collection of valid data is to solve technical issues with sampling and data extraction.(1) In the routine data collection performed during our studies, several technical issues were experienced by different actors in the data collection process. At the patient level, the patients could provide their PROMs via different routes in the EHRs: some at home using online portals and some at the physical therapy practice during the treatment. Previous research suggested that supportive EHR-systems should be developed to enable patients to fill in their PROMs via online portals at home to increase the amount and reliability of the data provided.(2) At the level of the physical therapists, some therapists did not register outcomes at all, and some selected versions of PROMs or physical performance measures included in the EHR were not connected to the national data registry, which led to a loss of data. At the level of software companies, a few used incorrect algorithms to calculate item scores with the result that the PROM scores in the national data registries could not be interpreted. Finally, at the level of the professional bodies in physical therapy, the standardisation procedures led to problems in their national data registries during the data collection by different EHR providers. Despite these challenges, a positive observation is that the amount of data collected was greater than for previous studies reported in Dutch physical therapy using the same registries.(3-6)

Methodological considerations

In this thesis, real-world observational outcomes of 72,226 treatment episodes for NSLBP and 4651 treatment episodes for COPD are presented. Although this seems a large amount, pre- and/or post-treatment measurements for the selected PROMs or performance measures were missing for a substantial proportion of the patients. For the analysis involving the patients with COPD, when the treatment episode had not ended, we used the last provided measurement. The proportion of adequate data fluctuated between 0.05 and 0.52 for patients with COPD, and between 0.07

and 0.33 for the different selected measurements for patients with NSLBP. Thus, it is still important to increase the proportion of patients with completed pre- and post-treatment measures. Without a second measurement after the baseline, the treatment outcomes (defined as the changes in outcomes at the end of treatment) cannot be estimated, and patients with missing data cannot be included in most of the multi-level analyses. As a result, the data of fewer patients are used to compare the differences in outcomes related to the treatment between practices, which may lead to selection bias. One method to search for potential selection bias is to compare the included patients (with pre- and post-treatment measurements) with patients lacking the pre- and post-treatment measurements, for instance to check whether there were any differences in treatment duration or patient characteristics. Previous research with patients with NSLBP used part of the same national registry database and judged that no selection bias was found between patient groups with and without missing data, based on a reliability analysis of a priori–formulated hypotheses, including a comparison of the patient characteristics of both groups. (7) It takes time and effort to develop routine data collection systems for PROMs in daily practice, as well as requiring effort to change behaviour and to develop the supporting systems.(8, 9)

A possible explanation for providing insufficient pre- and post-treatment measurements could be that patients with short treatment trajectories do not consider the completion of these measurements to be useful. We hypothesise that these patients are less motivated to complete post-treatment scores because they no longer have any complaints and/or the treatment trajectory is already finished. This hypothesis is strengthened by the results of a previous study comparing patients with and without pre- and post-treatment PROM scores, which revealed that the patients lacking these measurements had received, on average, 35% fewer treatment episodes than those who did complete them.(7) In these cases, a solution could be that patients do not need to score all PROM items but only report their recovery, or alternatively that the PROMS outcome could be coupled to the Global Perceived Effect scale.

Another explanation for the low response rates could be that patients experience survey fatigue, which can lead to a decline in survey completion over time.(10) Other studies showed that a low response rate was associated with older age, chronicity, comorbidities, questionnaire length, item relevance or perceptions of response burden.(11, 12) An important aspect for increasing patient awareness of the usefulness of providing PROMs is to use the provided outcomes as an evaluation of the treatment course.(13-15) Nonetheless, despite the potentially perceived response burden, Atkinson et al. (12) found that a growing number of patients with cancer are willing to self-report their experiences for themselves and others, and suggested that it should be possible to increase responses across a broad range of patients with different diseases.

Recommendations for practice and research

Future efforts should focus on solving the technical issues in data collection, with end-to-end validations to search for bugs or other technical problems. In end-to-end validations, standardised dummy treatment episodes, including PROMs and physical performance measures, are entered into the EHRs and sent to the national registries. The dummy treatment episodes are then checked in the registries for completeness, and potential technical issues can be found and solved in collaboration with the responsible actors.

Another aspect for increasing the validity and reliability of the outcome measurements is to consider the method used to collect patient outcomes. Current Dutch practice is that physical therapists are responsible for providing outcome data in the national data registries via their EHR, while most collected outcomes in the standard sets are PROMs completed by patients. A possible future initiative could be to create a system in which patients are responsible for providing data without involvement of their physical therapists. This system needs to be understandable and feasible for all patients, e.g., via an app, website, or email, but should stay connected to the EHR data. Several studies demonstrated that the routine use of PROMs is acceptable and feasible for patients, who expressed a preference for an electronic mode of administration.(16-18)

Furthermore, tailored implementation interventions could help increase data collection. For example, Eilayyan et al.(15) identified several barriers for the use of PROMs in primary care clinical settings, including a lack of skills, beliefs about the consequences, and the environmental context, so they developed a theory-based knowledge-translation intervention to facilitate the use of PROMs based on the identified barriers. Strategies such as this can increase the routine completion of outcome measures and thus increase the validity of data used to develop quality indicators.

Development of outcome-based quality indicators

Interpretation of the results

The inclusion of stakeholders (patients and representatives of patient associations, physical therapists, policy-makers of the professional bodies and health insurers) in the development process of this thesis was essential for ensuring that the core set of outcome-based quality indicators was feasible for daily practice. Without the collaboration and acceptance of stakeholders, our studies would not have succeeded. In the steering committee meetings involving multiple stakeholder representatives held throughout the project, we discussed the views and perspectives of stakeholders regarding the value and implementation of outcome-based quality indicators for Dutch physical therapy. Furthermore, we used the national data registries of the professional bodies in Dutch physical therapy for our data collection,

and Dutch health insurers encouraged the participation of physical therapists in our studies. The patient associations monitored the process in the stakeholder meetings. Other research has also underlined the importance of stakeholder engagement in designing implementation strategies for quality-improvement purposes, (19-22) and for overcoming barriers in the implementation of PROMs in daily practice.(11, 23)

We reached a consensus of multiple stakeholders (patients and patient associations, and purchasers and providers of care) regarding the standard sets of outcome domains and associated measures. The final selection of the core sets of outcome-based quality indicators for NSLBP and COPD based on these standard sets was conducted by physical therapists and senior researchers alone, enabling a specific focus on quality-improvement purposes by providing feedback of the patient outcomes in physical therapy practice. Nonetheless, as pointed out in the previous paragraph, patient associations and health insurers were informed and gave consent during the routinely organised meetings.

Methodological considerations

In developing quality indicators, adjusting the outcomes for case-mix variables that influence the outcomes of delivered care is highly important for the comparability and discriminative ability of health outcomes between physical therapy practices. In our studies, we were only able to adjust for a limited number of patient characteristics: age, gender, chronicity and the baseline score of the outcome of interest. Other patient level case-mix adjusters may also be relevant in physical therapy practice, such as socioeconomic status, ethnicity, smoking history, comorbidities, work status or psychosocial factors (poor social support, anxiety, depression and catastrophising).(24-26) By including more relevant patient-level case-mix variables in the multi-level analysis, the residual variance is expected to decrease, and thus the comparability between physical therapy practices would increase.(27) Future data collection should try to include more meaningful patient characteristics that can be used as case-mix variables; however, the administrative burden of patients and physical therapists must also be taken into account. One possible solution would be to develop a questionnaire to collect case-mix variables that can be administered via the EHR by patients in their own homes. Also, some case-mix variables could be derived from standardised procedures in the EHR without effort for patients or physical therapists, such as estimating socioeconomic status based on the postal code of patients.

Another possibility for dealing with differences in the characteristics of patients when developing quality indicators is the stratification of patients into subgroups. The value of stratifying heterogenous patient populations for targeted treatment options and to increase the comparability of treatment outcomes between stratified patients is well documented.(28-36) From the start of this thesis, we were aware of the added

value of stratification and we aimed to use implement this approach in the multi-level analyses to increase the comparability of the aggregated treatment outcomes between physical therapists. The physical therapists in our interviews described in chapter 6 also underlined the value of stratifying patients with COPD into subgroups based on the burden of disease. This is especially important when focussing on a heterogeneous patient population such as those with COPD. Unfortunately, we did not have sufficient data to stratify patients from either population (those with NSLBP or COPD) into subgroups for a multi-level analysis at the physical therapy practice level.

At the total patient population level however, we were able to stratify patients with NSLBP and identified differences between the subgroups. For these patients, the experienced pain and physical functioning were significantly different between subgroups based on a stratification by prognostic factors using the STarT Back Screening Tool (SBT). Still, previous research concluded that prognostic screening instruments, such as the SBT, scored poorly at assigning higher risk scores to individuals who develop chronic pain than to those who will not.(31) A recent publication by Bier et al. (37) increased the predictive validity of the Dutch SBT by incorporating the duration of the complaints and changing cut-off scores of the screening tool, which may potentially increase the usability of the SBT as a stratification tool.

Recommendations for research and practice

In our opinion, future research should focus on the further development of valid stratification tools or models. One example would be the estimation of the validity of the recently developed profiling system for patients with COPD, used to allocate patients into subgroups for exercise-based care, presented as the “Dutch model”. (30) This Dutch Model has the aim of assigning the right patient with COPD to the right type of exercise-based care at the right moment.(30) Through the use of valid stratification tools, the comparability and discriminability of quality indicators between physical therapy practices will increase, and will therefore be more useful for quality-improvement initiatives or public reporting.

Using outcomes in clinical decision-making

Interpretation of the results

As stated in our general introduction, monitoring quality of care is shifting from evaluating processes of care towards evaluating outcomes of care. A benefit of this shift is the reduction of the experienced administrative burden on physical therapists in administrating processes of care. When physical therapists collect a minimal standard set of outcomes, of which most are PROMs that can be provided via online portals by patients in their EHR, the remaining administration will be less time-consuming during the treatment. Although some EHR systems already make

it possible for patients to complete their PROMs online at home, there is still room for improvement to actually motivate patients to do so. During the interviews with physical therapists in our pilot study described in chapter 6, the therapists typically underlined the value of using PROMs in their communication with patients. This is supported by other research into the use of PROMs in clinical practice, which can encourage physical therapists to motivate their patients to complete their PROMs online.(14, 38-40) This collection of meaningful outcomes for patients that can be used in the interaction between patients and physical therapists can also be used to evaluate quality of care. A precondition seems to be that the administrative burden of physical therapists is as low as possible by providing PROMs for online completion by patients at home, with the registration of a limited number of physical performance measures during the practice visit.

Methodological considerations

A central topic in the interaction of physical therapists and patients during clinical decision-making is evaluating the course of recovery. In our studies we only used cross-sectional outcome data for the development of quality indicators, such as post-treatment scores and change scores. This raises the possibility of using outcome data routinely collected during the treatment episode to predict the course of recovery through a 'people like me' approach to support person-centredness and individualised care.(41) Using the 'people like me' approach, the course of recovery can be predicted using data from similar patients based on their characteristics and the recovery curve, and can be used for shared decision-making and goal setting by an individual patient and their physical therapist. Currently, Dutch researchers and physical therapists are using this system in patients with intermittent claudication.(42)

When reflecting on the 'people like me' approach and the populations in our studies, we noted that patients with COPD have a greater potential to use this approach because their standard set included physical performance measures, while the standard set for patients with NSLBP only included PROMs. As reported by A. J. Kittelson et al.,(41) "measures of physical performance may be more sensitive to change (thus potentially more useful for monitoring progress) and provide a different picture of patient functioning than is captured with PROMs". Nonetheless, PROMs report on patient perceptions of their daily functioning and participation in society, which are not directly assessed by physical performance measures. Furthermore, we found that the PROMs for patients with NSLBP were able to discriminate between physical therapy practices.(43) Another possibility is to combine PROMs with physical performance measures to explore the association of these trajectories with one-year outcomes, as Harmelink et al. demonstrated.(44) They combined the Timed Up and Go (TUG) a physical performance measure with the Knee injury and Osteoarthritis Outcome Score-Activities of Daily Living (KOOS-ADL) a PROM for measuring physical functioning.(44)

Recommendations for research and practice

A notable benefit in shifting towards outcomes of care is that the meaningful outcomes of patients as presented in our standard sets can also be explored in the interactions between patients and physical therapists, e.g., in goal setting and shared decision-making.(3, 45-48) The usefulness of using routinely collected outcomes for goal setting and shared decision-making is well documented.(14, 40, 42, 49, 50) A recent Cochrane review by Gibbons et al. (50) stated that "Despite the mixed certainty of the evidence, mainly due to issues with blinding and concealment which are difficult to overcome in trials of complex interventions that include feedback elements, the data suggest that routine use of PROM feedback in clinical practice could thus improve the quality of health care". Physical therapists who treated patients with COPD underlined the value of the standard set in their interaction with patients; for example, the physical therapists stated that patients were motivated to complete the measurement instruments in the standard set to monitor their own results over time. Furthermore, the physical therapists highlighted that the standard sets guide their treatment plan and, if possible, improve the treatment results.

Using quality indicators for internal quality improvement

Interpretations of the results

We developed core sets of outcome-based quality indicators that are able to discriminate patient outcomes between physical therapy practices, which can be used as a basis to search for explanations of differences in delivered care. Quality improvement can be achieved by reflecting on patient outcomes with peers and learning from each other for future treatments.(51, 52)

As an illustration, figure 1 presents a graphic visualisation of a quality indicator: pre- and post-treatment score changes in experienced pain, as measured using the Numeric Pain Rating Scale (NPRS), were sampled in 22,740 patients with NSLBP. The data were clustered within 93 practices and adjusted for case-mix variables. Roughly one third of the practices scored less than average, one third were average, and one third were above average, based on the confidence intervals (95% CI) of each practice. Practices that are presented in green registered higher changes scores than the practices in purple and blue. This can be an indication of potential differences in the quality of delivered care between practices. The next step would be to search for explanations of those differences in outcomes within the processes of delivered care: what was the content of the delivered care? Are the patient populations substantially different from the benchmark data? Are the differences in change scores clinically relevant? Hence, the interpretation of the differences in delivered care must be clarified to identify whether these differences are indeed related to quality of care. The clarification of outcomes is, in our opinion, most effective when physical therapists collaborate with their peers. In the interpretation of the results

with peers, the opportunity arises to compare, discuss, and learn from each other through clinical reasoning processes.(51-53) In our opinion interpreting outcomes and discussing with peers are the mechanisms through which quality improvement can be achieved. In the following paragraphs, we provide some suggestions for using quality indicators to evaluate quality of care.

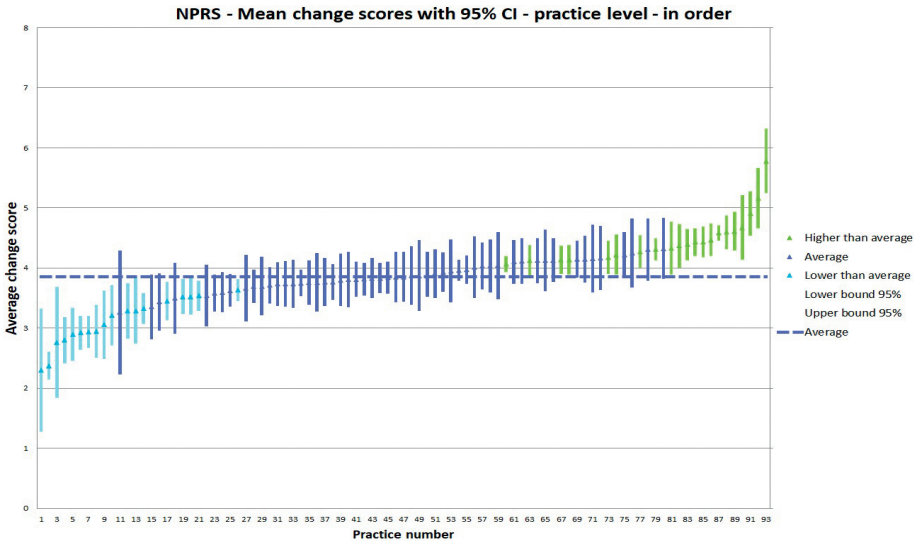


Figure 1 Monitoring the outcome of a score change in pain intensity measured using the NPRS

Physical therapy practices can use the core sets to monitor and compare treatment outcomes in daily practice in order to evaluate and enhance quality of care. In this thesis, we performed a practice test, comparing the outcomes of cross-sectionally measured outcome-based quality indicators between individual physical therapy practices. To develop a full picture of the validity and usefulness of the core sets, further research will be needed to determine which quality indicators are measured over time. Quality indicator scores could then be used for longitudinal evaluation and the monitoring of quality improvement within physical therapy practices.

Methodological considerations

A possible challenge in the usefulness of the quality indicators within physical therapy practices is the reliability of comparing the outcomes of patient populations with a low prevalence. Most physical therapists within a practice have different target populations based on their expertise, which could pose a challenge for the treatment of populations of patients with less common conditions, such as those with elbow or hand conditions. A solution could be to cluster these diagnoses within musculoskeletal disorders and compare patient outcomes on general indicators such

as physical functioning or pain, which would enhance comparability at the practice level. When physical therapists use the aggregated outcomes of their patients to compare and learn from the treatment outcomes of their peers, they may also be interested in more patient-specific measures.

Recommendations for research and practice

A variety of recommendations regarding the use of quality indicators for internal quality improvement in future practice and research can be derived from the different studies in this thesis. Despite the advancements in the routine use of PROMs to evaluate quality of care, doubts remain about the quantifiable benefits of implementing PROMs in current clinical practice.(11, 54, 55) While the potential impact of PROMs on patient–clinician communication has been well investigated, knowledge about the impact of routine outcome measurement on improving patient outcomes such as quality of life is still limited.(14, 50, 56-59). We think that feedback on (aggregated) patient outcomes can be beneficial for improving quality of care in several aspects of health care.

An important requirement in benchmarking and comparing outcomes between peers is an adequate data infrastructure. The Association for Quality in Physical Therapy (SKF) and the Royal Dutch Society for Physical Therapy (KNGF) have developed dashboards for physical therapists and practice owners to receive feedback on outcomes in comparison with benchmark data.(60, 61) Through the dashboard, physical therapists also receive feedback on the proportion of patients with pre- and post-treatment measures (processes of care) in the database. With this feedback system, the amount of data provided in the database is transparent for each individual physical therapist. When sufficient data are provided, including case-mix variables, the dashboard can be used as a learning tool for interpreting the differences between peers.

The management of physical therapy practices can use routinely collected outcomes to monitor and compare differences between their employees or benchmark data, or they can use the outcomes to conduct a plan-do-study-act (PDSA) cycle to provide a structured iterative testing and evaluation of changes in their care processes and outcomes to improve quality of care.(62) When using the PDSA cycle, physical therapists can define action plans based on their own specific learning goals. The quality indicators can play a key role in this process, thus leading to improvement in the care of patients with NSLBP and COPD.

The effectiveness of the feedback on (aggregated) outcomes for quality improvement strongly depends on the method that is used.(63) Feedback is more likely to be accepted when it comes from a reliable source, such as a supervisor or colleague. (63) To our perspective, one of the most promising future efforts for quality

improvement in daily practice is using real-world collected outcomes as a learning tool in peer assessment meetings. In such meetings, professionals reflect on their own performance, critically appraise their peers' performance and give constructive feedback.(52, 53) When physical therapists receive feedback using combined outcome measures, they may gain more insight into the quality of their performance; however, our hypothesis is that feedback on health outcomes alone will not lead to quality improvement. Instead, professionals need to interpret the data and translate this information into meaningful actions or use their outcomes in clinical reasoning discussions. The standard sets of outcomes and outcome-based quality indicators can be used as a learning tool during the peer meetings, facilitating the comparison of outcomes and the discussion of alternatives for future treatments.

A prerequisite is that the members of a peer group feel safe enough to share and discuss routinely collected data.(51) A safe environment is encouraged by including well-trained coaches in each peer group meeting. The coach needs to be educated in conducting structured meetings, with an emphasis on the importance of creating a safe environment.(51, 53) Currently, conducting peer assessment meetings is mandatory for members of the SKF and for some physical therapy practices of the KNGF. Both the KNGF and SKF recently developed manuals that can guide peer assessment meetings (with a coach) in using aggregated outcomes to compare performance and promote quality improvement.(64, 65) The next step is to investigate the impact of these meetings on the quality of delivered care.

Using quality indicators for external transparency

Interpretation of the results

A key conclusion of our interviews with physical therapists was that they underlined the value of transparency regarding their treatment outcomes for quality-improvement purposes. They valued providing this transparency to policy-makers and in the (anonymised) publication of outcomes, rather than making their outcomes transparent for health insurers. One important reason not to provide fully transparent data yet was, as described in previous paragraphs, concern about the quality of the data, which further highlights the need to increase the proportion of patients for whom valid data is collected. An example of the need to increase the validity of the data is the stratification of patients with COPD into subgroups based on the burden of disease, or the grouping of patients with NSLBP based on their prognosis profile. Furthermore, physical therapists were concerned about the consequences for reimbursement by health insurers. There were also doubts about the knowledge and skills of patients to interpret the aggregated outcomes when choosing their healthcare provider.

Methodological considerations

The manipulation or 'gaming' of data can be an unintended consequence of using outcome-based quality indicators for public reporting, particularly in pay-for-performance initiatives.(13) In our perspective, gaming means that providers manipulate outcomes because of a fear of negative consequences or to increase the positive consequences of their performance. This was suggested during our interviews with physical therapists. An example of a negative consequence is the loss of reimbursement of physical treatments if norm values of quality indicators are not achieved. With this top-down initiative, the goal of improving quality of care may instead result in physical therapists simply trying to reach the norm value by gaming their data.

A review that described the results of pay-for-performance initiatives in 14 countries indeed concluded that it remains unclear whether quality of care was improved by such schemes;(66) however, the same review suggested that public reporting was positively associated with quality improvements, and may represent a suitable alternative to pay-for-performance initiatives with an even stronger financial incentive.(66-70) In our studies, physical therapists supported the use of making outcomes transparent at the practice level, but anonymously using numbers, such as is shown in figure 1. We considered providing aggregated outcome-based quality indicators anonymously as a first step in working towards full transparency, enabling physical therapists to compare the differences between their outcomes and those of other practices without violating their privacy. We think this is critical for maintaining the value of outcome-based quality indicators as learning tools. Again, creating a safe environment for physical therapists is key for such implementation initiatives.(52) The next steps towards external transparency in the future can only be made together with the physical therapists themselves.

We believe that gaming is not common in the current Dutch physical therapy practice, as the collection of patient outcomes occurs in a safe environment. This hypothesis is supported by previous research into Dutch primary care physical therapy comparing EHR data with those from surveys, which showed that the completeness of the processes of care was above 90 percent for all indicators in both the survey data and EHR data.(5) Furthermore, reflecting on our own data, we observed a large variation in outcomes between physical therapy practices. This indicates that there is room for improvement by searching for explanations with peers when outcomes differ. Furthermore, in (focus group) interviews, the physical therapists displayed an intrinsic motivation to collect outcomes for learning purposes: physiotherapists are curious to know how they perform.

Recommendations for research, policy and practice

Aggregated routinely collected outcome data can be used to support future policy decisions in health care. This can be done by evaluating the value of the delivered primary care physical therapy treatments at the population level, but also by comparing the value of a given physical therapy treatment with other (more expensive) treatments, which can be used to develop policies surrounding the substitution of care. Substitution of care can be defined as 'the continual regrouping of resources across and within care settings to exploit the best and least costly solution in the face of changing needs and demands'.(71) One example would be the shift from specialised hospital care and unnecessary hospital consultations, which are generally more expensive, to less expensive primary care (such as physical therapy care),(72) a change which is supported by the Dutch minister of Health, Welfare and Sports.(73)

The standard sets developed in this thesis are used in Dutch daily physical therapy practice for quality improvement, and are mandatory for all physical therapy practices participating in SKF. By collecting outcomes from the standard sets, SKF practices can be transparent to external stakeholders, such as patients and health insurers, demonstrating that they are using data collection in combination with the peer assessment meetings to improve the quality of care. Currently, Dutch health insurers receive policy reports on the process of data collection at each SKF practice, but not the outcomes of the measures. As discussed, we believe health insurers could encourage the use of data collection, however differences in outcomes would need to be interpreted and discussed by the physical therapists themselves in a safe environment.

New initiatives have also been introduced in the domain of external transparency of quality of care. Currently, health insurers base their reimbursement policy for Dutch physical therapy practices on the treatment index as a proxy for quality of the delivered care. In chapter 1, we explained that the treatment index gives insight into the number of treatment sessions per patient in a physical therapy practice compared with the overall mean number of treatment sessions per patient in all practices. The value of this treatment index as an instrument for controlling healthcare costs and quality is not without controversy, as it only focusses on the number of treatment sessions in terms of costs and lacks a component to evaluate the beneficial outcomes of the provided care,(74) e.g., fewer treatment sessions could have less of a benefit. The SKF is exploring the feasibility of an outcome index using real-world data based on the Value-Based Health Care principles of Porter and Teisberg.(75) Value-Based Health Care aims to achieve high value for patients, with value defined as the health outcomes relative to the costs for achieving these outcomes.(75, 76) In the outcome index, the health outcomes are expressed as changes in pain and/or physical functioning, while the costs are expressed in the number of treatments and

the duration of the treatment episode. This is an example of how routinely collected patient outcomes may be used to evaluate quality of care and external transparency.

Potential future initiatives can use outcomes of delivered care for public reporting, e.g., by presenting outcomes at the population level to gain information about the overall effect experienced by the patients who were treated in the Netherlands. Through these outcomes, the value of Dutch physical therapy practice for patients can become transparent. Important prerequisites for the use of outcomes for external transparency at the physical therapist level include ensuring the validity of the data, solving technical issues and creating a system through which patients can provide patient-reported data without the involvement of their physical therapists, while ensuring the data are connected to the EHR.

OVERALL CONCLUSIONS

We conclude that the standardised routine data collection of patient-reported outcomes of care combined with performance measures can play a key role for quality improvement in physical therapy practice. An important prerequisite in using outcomes for quality improvement is that they are comparable and able to discriminate between physical therapists or physical therapy practices. In this thesis, we developed standard sets and core sets of outcome-based quality indicators for physical therapy patients with NSLBP and for patients with COPD. The sets were accepted by stakeholders as having added value for the interactions between patients and physical therapists, for quality-improvement purposes, and for public reporting at the population level.

Future efforts should focus on increasing the validity and reliability of the data collected by the participating physical therapists and patients, including relevant patient characteristics that can be used to increase the comparability of the outcome data by taking patient-mix variables into account. Patients should be able to complete their PROMs using an understandable user interface in a safe environment without interference from physical therapists. Physical therapists should receive tools to use outcome-based quality indicators for learning purposes to improve quality of care in a cyclical process. The next step towards external transparency in the future should be made by the stakeholders, including physical therapists themselves, in a collaborative approach.

REFERENCES

1. Meerhoff GA, van Dulmen SA, Maas MJM, Heijblom K, Nijhuis-van der Sanden MWG, Van der Wees PJ. Development and Evaluation of an Implementation Strategy for Collecting Data in a National Registry and the Use of Patient-Reported Outcome Measures in Physical Therapist Practices: Quality Improvement Study. Physical Therapy. 2017;97(8):837-51.
2. Meerhoff GA, Dulmen SAv, Cruijsberg JK, Sanden MWGN-vd, Wees PJVd. Which Factors Influence the Use of Patient-Reported Outcome Measures in Dutch Physiotherapy Practice? A Cross-Sectional Study. Physiotherapy Canada. 2020;72(1):63-70.
3. van Dulmen SA, van der Wees PJ, Bart Staal J, Braspenning JCC, Nijhuis-van der Sanden MWG. Patient reported outcome measures (PROMs) for goalsetting and outcome measurement in primary care physiotherapy, an explorative field study. Physiotherapy. 2017;103(1):66-72.
4. Scholte M, Neeleman-van der Steen CW, Hendriks EJ, Nijhuis-van der Sanden MW, Braspenning J. Evaluating quality indicators for physical therapy in primary care. International journal for quality in health care : journal of the International Society for Quality in Health Care. 2014;26(3):261-70.
5. Scholte M, van Dulmen SA, Neeleman-Van der Steen CW, van der Wees PJ, Nijhuis-van der Sanden MW, Braspenning J. Data extraction from electronic health records (EHRs) for quality measurement of the physical therapy process: comparison between EHR data and survey data. BMC medical informatics and decision making. 2016;16(1):141.
6. Scholte M, Calsbeek H, Nijhuis-van der Sanden MW, Braspenning J. Quality of physical therapy from a patient's perspective; factor analysis on web-based survey data revealed three dimensions on patient experiences with physical therapy. BMC health services research. 2014;14:266.
7. Meerhoff GA, Verburg AC, Schapendonk RM, Cruijsberg J, Nijhuis-van der Sanden MWG, van Dulmen SA, et al. Reliability, validity and discriminability of patient reported outcomes for non-specific low back pain in a nationwide physical therapy registry: A retrospective observational cohort study. PLoS One. 2021;16(6):e0251892.
8. Cox CE, Wysham NG, Kamal AH, Jones DM, Cass B, Tobin M, et al. Usability Testing of an Electronic Patient-Reported Outcome System for Survivors of Critical Illness. Am J Crit Care. 2016;25(4):340-9.
9. Steele Gray C, Gill A, Khan AI, Hans PK, Kuluski K, Cott C. The Electronic Patient Reported Outcome Tool: Testing Usability and Feasibility of a Mobile App and Portal to Support Care for Patients With Complex Chronic Disease and Disability in Primary Care Settings. JMIR Mhealth Uhealth. 2016;4(2):e58.
10. Buergy D, Siefert V, Neumaier C, Ganslandt T, Sperk E, Blessing M, et al. Prospective trial on telemonitoring of geriatric cancer patients using handheld devices. Strahlenther Onkol. 2020;196(3):205-12.
11. Aiyegbusi OL, Kyte D, Cockwell P, Marshall T, Dutton M, Walmsley-Allen N, et al. Patient and Clinician Perspectives on Electronic Patient-Reported Outcome Measures in the Management of Advanced CKD: A Qualitative Study. Am J Kidney Dis. 2019;74(2):167-78.
12. Atkinson TM, Schwartz CE, Goldstein L, Garcia I, Storfer DF, Li Y, et al. Perceptions of Response Burden Associated with Completion of Patient-Reported Outcome Assessments in Oncology. Value Health. 2019;22(2):225-30.

13. Greenhalgh J, Dalkin S, Gooding K, Gibbons E, Wright J, Meads D, et al. Health Services and Delivery Research. Functionality and feedback: a realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care. Southampton (UK): NIHR Journals Library; 2017.

14. Greenhalgh J, Gooding K, Gibbons E, Dalkin S, Wright J, Valderas J, et al. How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. Journal of Patient-Reported Outcomes. 2018;2(1):42.

15. Eilayyan O, Visca R, Zidarov D, Ware P, Bussières A, Ahmed S. Developing theory-informed knowledge translation strategies to facilitate the use of patient-reported outcome measures in interdisciplinary low back pain clinical practices in Quebec: mixed methods study. BMC health services research. 2020;20(1):789.

16. Schick-Makaroff K, Molzahn A. Brief communication: patient satisfaction with the use of tablet computers: a pilot study in two outpatient home dialysis clinics. Canadian Journal of Kidney Health and Disease. 2014;1(1):22.

17. Dumais KM, Dias N, Khurana L, Gary ST, Witherspoon B, Evans CJ, et al. Preferences for Use and Design of Electronic Patient-Reported Outcomes in Patients with Chronic Obstructive Pulmonary Disease. Patient. 2019;12(6):621-9.

18. Schick-Makaroff K, Molzahn AE. Evaluation of real-time use of electronic patient-reported outcome data by nurses with patients in home dialysis clinics. BMC health services research [Internet]. 2017 2017/06//; 17(1):[439 p.]. Available from: http://europepmc.org/abstract/MED/28651528

19. Hamilton AB, Brunner J, Cain C, Chuang E, Luger TM, Canelo I, et al. Engaging multilevel stakeholders in an implementation trial of evidence-based quality improvement in VA women's health primary care. Translational behavioral medicine. 2017;7(3):478-85.

20. Hamilton AB, Yano EM. The importance of symbolic and engaged participation in evidence-based quality improvement in a complex integrated healthcare system: response to "The science of stakeholder engagement in research". Translational behavioral medicine. 2017;7(3):492-4.

21. Goodman MS, Sanders Thompson VL. The science of stakeholder engagement in research: classification, implementation, and evaluation. Translational behavioral medicine. 2017;7(3):486-91.

22. Campbell SM, Braspenning J, Hutchinson A, Marshall MN. Research methods used in developing and applying quality indicators in primary care. BMJ. 2003;326(7393):816-9.

23. Stover AM, Tompkins Stricker C, Hammelef K, Henson S, Carr P, Jansen J, et al. Using Stakeholder Engagement to Overcome Barriers to Implementing Patient-reported Outcomes (PROs) in Cancer Care Delivery: Approaches From 3 Prospective Studies. Med Care. 2019;57 Suppl 5 Suppl 1:S92-s9.

24. Knezevic NN, Candido KD, Vlaeyen JWS, Van Zundert J, Cohen SP. Low back pain. The Lancet. 2021;398(10294):78-92.

25. GOLD. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. Fontana. Global Initiative for Chronic Obstructive Lung Disease (GOLD). 2020.

26. Gibson GJ, Loddenkemper R, Lundbäck B, Sibille Y. Respiratory health and disease in Europe: the new European Lung White Book. European Respiratory Journal. 2013;42(3):559-63.

27. Twisk JWR. Applied Mixed Model Analysis: A Practical Guide. 2 ed. Cambridge: Cambridge University Press; 2019.

28. Bier JD, Sandee-Geurts JJW, Ostelo R, Koes BW, Verhagen AP. Can Primary Care for Back and/or Neck Pain in the Netherlands Benefit From Stratification for Risk Groups According to the STarT Back Tool Classification? Archives of physical medicine and rehabilitation. 2018;99(1):65-71.

29. Foster NE, Mullis R, Hill JC, Lewis M, Whitehurst DG, Doyle C, et al. Effect of stratified care for low back pain in family practice (IMPaCT Back): a prospective population-based sequential comparison. Annals of family medicine. 2014;12(2):102-11.

30. Spruit MA, Van't Hul A, Vreeken HL, Beekman E, Post MHT, Meerhoff GA, et al. Profiling of Patients with COPD for Adequate Referral to Exercise-Based Care: The Dutch Model. Sports Medicine. 2020;50(8):1421-9.

31. Karran EL, McAuley JH, Traeger AC, Hillier SL, Grabherr L, Russek LN, et al. Can screening instruments accurately determine poor outcome risk in adults with recent onset low back pain? A systematic review and meta-analysis. BMC medicine. 2017;15(1):13.

32. Foster NE, Hill JC, O'Sullivan P, Hancock M. Stratified models of care. Best practice & research Clinical rheumatology. 2013;27(5):649-61.

33. Alrwaily M, Timko M, Schneider M, Stevans J, Bise C, Hariharan K, et al. Treatment-Based Classification System for Low Back Pain: Revision and Update. Phys Ther. 2016;96(7):1057-66.

34. Fritz JM. Clinical Prediction Rules in Physical Therapy: Coming of Age? Journal of Orthopaedic & Sports Physical Therapy. 2009;39(3):159-61.

35. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. BMJ : British Medical Journal. 2013;346:e5595.

36. Hingorani AD, Windt DAvd, Riley RD, Abrams K, Moons KGM, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. BMJ : British Medical Journal. 2013;346:e5793.

37. Bier JD, Kuijer MR, Jong A, Verhagen A. Improving the Predictive Validity of the Dutch STarT Back Tool. Physical Therapy. 2021.

38. Greenhalgh J. The applications of PROs in clinical practice: what are they, do they work, and why? Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2009;18(1):115-23.

39. Santana M-J, Feeny D. Framework to assess the effects of using patient-reported outcome measures in chronic care management. Quality of Life Research. 2014;23(5):1505-13.

40. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. BMC health services research. 2013;13:211.

41. Kittelson AJ, Hoogeboom TJ, Schenkman M, Stevens-Lapsley JE, van Meeteren NLU. Person-Centered Care and Physical Therapy: A "People-Like-Me" Approach. Phys Ther. 2020;100(1):99-106.

42. Sinnige A, Spruijt S, Saes M, Van der Wees PJ, Hoogeboom TJ, Teijink JAW. Using a Learning Health System to Improve Physical Therapy Care for Patients with Intermittent Claudication: Lessons Learned from the ClaudicatioNet Quality System. Physical Therapy. 2021.

43. Verburg AC, van Dulmen SA, Kiers H, Nijhuis-van der Sanden MWG, van der Wees PJ. Patient-Reported Outcome-Based Quality Indicators in Dutch Primary Care Physical Therapy for Patients With Nonspecific Low Back Pain: A Cohort Study. Phys Ther. 2021.

44. Harmelink KEM, Dandis R, der Van der Wees Pj PJ, Zeegers A, der Sanden MWN, Staal JB. Recovery trajectories over six weeks in patients selected for a high-intensity physiotherapy program after Total knee Arthroplasty: a latent class analysis. BMC musculoskeletal disorders. 2021;22(1):179.

45. Meerhoff GA, van Dulmen SA, Maas MJM, Bakker-Jacobs A, Nijhuis-Van der Sanden MWG, van der Wees PJ. Exploring the perspective of patients with musculoskeletal health problems in primary care on the use of patient-reported outcome measures to stimulate quality improvement in physiotherapist practice; a qualitative study. Physiotherapy theory and practice. 2019:1-12.

46. Stevens A, Koke A, van der Weijden T, Beurskens A. The development of a patient-specific method for physiotherapy goal setting: a user-centered design. Disability and rehabilitation. 2018;40(17):2048-55.

47. Verburg AC, van Dulmen SA, Kiers H, Ypinga JH, Nijhuis-van der Sanden MW, van der Wees PJ. Development of a Standard Set of Outcome Domains and Proposed Measures for Chronic Obstructive Pulmonary Disease in Primary Care Physical Therapy Practice in the Netherlands: a Modified RAND/UCLA Appropriateness Method. International journal of chronic obstructive pulmonary disease. 2019;14:2649-61.

48. Verburg AC, van Dulmen SA, Kiers H, Nijhuis-van der Sanden MWG, van der Wees PJ. Development of a standard set of outcome measures for non-specific low back pain in Dutch primary care physiotherapy practices: a Delphi study. European Spine Journal. 2019;28(7):1550-64.

49. Stevens A, Koke A, van der Weijden T, Beurskens A. Ready for goal setting? Process evaluation of a patient-specific goal-setting method in physiotherapy. BMC health services research. 2017;17(1):618.

50. Gibbons C, Porter I, Gonçalves-Bradley DC, Stoilov S, Ricci-Cabello I, Tsangaris E, et al. Routine provision of feedback from patient-reported outcome measurements to healthcare providers and patients in clinical practice. The Cochrane database of systematic reviews. 2021;10(10):Cd011589.

51. Maas MJ, Nijhuis-van der Sanden MW, Driehuis F, Heerkens YF, van der Vleuten CP, van der Wees PJ. Feasibility of peer assessment and clinical audit to self-regulate the quality of physiotherapy services: a mixed methods study. BMJ open. 2017;7(2):e013726.

52. Maas MJM, Driehuis F, Meerhoff GA, Heerkens YF, van der Vleuten CPM, Nijhuis-van der Sanden MWG, et al. Impact of Self- and Peer Assessment on the Clinical Performance of Physiotherapists in Primary Care: A Cohort Study. Physiother Can. 2018;70(4):393-401.

53. Maas MJ, van der Wees PJ, Braam C, Koetsenruijter J, Heerkens YF, van der Vleuten CP, et al. An innovative peer assessment approach to enhance guideline adherence in physical therapy: single-masked, cluster-randomized controlled trial. Phys Ther. 2015;95(4):600-12.

54. Aiyegbusi OL, Isa F, Kyte D, Pankhurst T, Kerecuk L, Ferguson J, et al. Patient and clinician opinions of patient reported outcome measures (PROMs) in the management of patients with rare diseases: a qualitative study. Health and quality of life outcomes. 2020;18(1):177.

55. Schick-Makaroff K, Molzahn A. Strategies to use tablet computers for collection of electronic patient-reported outcomes. Health and quality of life outcomes. 2015;13(1):2.

56. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring Quality of Life in Routine Oncology Practice Improves Communication and Patient Well-Being: A Randomized Controlled Trial. Journal of Clinical Oncology. 2004;22(4):714-24.

57. Krawczyk M, Sawatzky R. Relational use of an electronic quality of life and practice support system in hospital palliative consult care: A pilot study. Palliative and Supportive Care. 2019;17(2):208-13.

58. Espallargues M, Valderas JM, Alonso J. Provision of feedback on perceived health status to health care professionals: a systematic review of its impact. Med Care. 2000;38(2):175-86.

59. Valderas JM, Kotzeva A, Espallargues M, Guyatt G, Ferrans CE, Halyard MY, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2008;17(2):179-93.

60. (KNGF) RDSfPT. National Database Physical Therapy (LDF) [Available from: https://www.kngf.nl/ldf.

61. The national data registry (LDK) of the Association for Quality in Physical Therapy (SKF) [Available from: https://www.keurmerkfysiotherapie.nl/praktijk/data/.

62. Taylor MJ, McNicholas C, Nicolay C, Darzi A, Bell D, Reed JE. Systematic review of the application of the plan-do-study-act method to improve quality in healthcare. BMJ Qual Saf. 2014;23(4):290-8.

63. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. The Cochrane database of systematic reviews. 2012(6):Cd000259.

64. Maas MJ, de Leur M, Wees PJvd, Kiers H. Peer review - meetinformatie gebruiken om te leren en verbeteren 2020 [Available from: https://www.keurmerkfysiotherapie.nl/wp-content/uploads/2020/06/peer-review-2020-thema-1.-data-1.0.pdf.

65. Maas MJ, van Poelgeest E, Weymans C, Wees PJvd. Module 3 feedback op proces en uitkomsten KNGF. 2018.

66. Milstein R, Schreyoegg J. Pay for performance in the inpatient sector: A review of 34 P4P programs in 14 OECD countries. Health Policy. 2016;120(10):1125-40.

67. Hospital Performance Reports: Impact On Quality, Market Share, And Reputation. Health Affairs. 2005;24(4):1150-60.

68. Trivedi AN, Nsa W, Hausmann LR, Lee JS, Ma A, Bratzler DW, et al. Quality and equity of care in U.S. hospitals. The New England journal of medicine. 2014;371(24):2298-308.

69. Williams SC, Schmaltz SP, Morton DJ, Koss RG, Loeb JM. Quality of Care in U.S. Hospitals as Reflected by Standardized Measures, 2002–2004. New England Journal of Medicine. 2005;353(3):255-64.

70. Romano PS, Marcin JP, Dai JJ, Yang XD, Kravitz RL, Rocke DM, et al. Impact of public reporting of coronary artery bypass graft surgery performance data on market share, mortality, and patient selection. Med Care. 2011;49(12):1118-25.

71. Warner M. Re-designing health services. Reducing the zone of delusion. London: The Nuffield Trust. 1997.

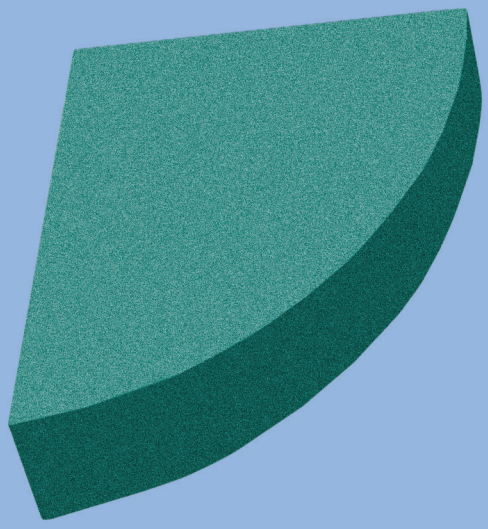
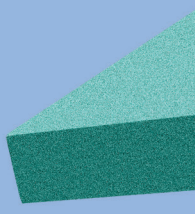
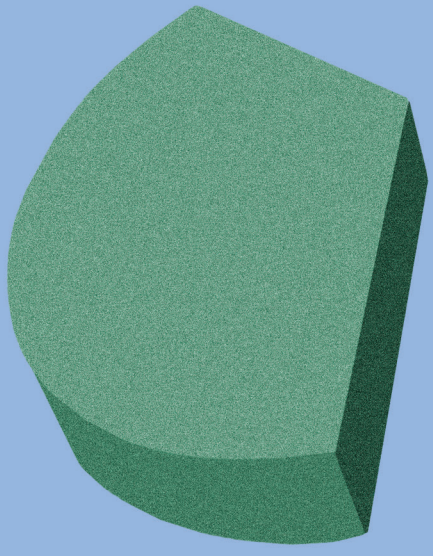
72. van Hoof SJM, Quanjel TCC, Kroese M, Spreeuwenberg MD, Ruwaard D. Substitution of outpatient hospital care with specialist care in the primary care setting: A systematic review on quality of care, health and costs. PLoS One. 2019;14(8):e0219957.

73. Westra D, Kroese M, Ruwaard D. [Substitution of care: what do we know, what should we know, and what should we do?]. Nederlands tijdschrift voor geneeskunde. 2017;161:D1354.

74. Skipr. Fysiotherapeut is bureaucratie beu [Available from: https://www.skipr.nl/blog/fysiotherapeut-is-bureaucratie-beu/.

75. Porter ME, E.O. T. Redefining health care: creating value-based competition on results.: Boston: Harvard Business School Press; 2006.

76. Porter ME. What is value in health care? The New England journal of medicine. 2010;363(26):2477-81.

Addendum

SUMMARY

The routine collection of treatment outcome data provides an opportunity to monitor, evaluate and improve quality of care in primary care physical therapy. These outcome data can be used at the individual level during the interactions between patients and professionals for goal setting, monitoring and evaluating treatment outcomes. When comparing aggregated outcomes at the group level, healthcare providers can reflect on their own performance and compare themselves with their peers. Furthermore, aggregated outcomes can be used for public reporting, e.g., as a supported tool enabling patients to choose a provider or for pay-for-performance initiatives.

The aim of this thesis was to develop, select and test core sets of outcome-based quality indicators in primary care physical therapy practice for patients with non-specific low back pain (NSLBP) and chronic obstructive pulmonary disease (COPD), based on consensus between stakeholders (patients and patient associations, physical therapists, policy-makers, researchers and health insurers). The core sets include relevant outcome domains and associated measures for monitoring and evaluating primary care physical therapy treatment, and their use is supported by routinely collected real-world data. Any comparison of meaningful outcomes between physical therapists needed to be valid, reliable and accepted as having added value by the stakeholders.

The first step in conducting this thesis was to determine which outcome domains (e.g., pain or physical functioning) should be measured for patients with NSLBP and COPD, and which measurements would be appropriate (e.g., Numeric Pain Rating Scale or Patient Specific Functional Scale) when evaluating physical therapy treatment. The two research questions for this step were therefore: 'What standard set of outcome domains and associated measures for patients with NSLBP can be developed in Dutch primary care physical therapy practice?' (chapter 2) and 'What standard set of outcome domains and associated measures for patients with COPD can be developed in Dutch primary care physical therapy practice?' (chapter 3). In chapters 2 and 3, we conducted a modified RAND/UCLA Delphi procedure to reach a consensus between stakeholders regarding the expected added value of the outcome domains and associated measures in clinical practice. Furthermore, we intended to include a proposal to stratify the patients into subgroups to enable the comparison of treatment outcomes between matched patients based on their characteristics. Both standard sets were developed to be useful for a) the interaction between the patient and the healthcare professional, b) internal quality improvement and c) the external transparency of primary care physical therapy practices.

In the second step, we aimed to define potential quality indicators based on the previously selected standard sets of outcome measures for patients with NSLBP

and COPD using the following research questions: 'Which potential outcome-based quality indicators for patients with NSLBP can be used for the selection of a core set based on the acceptance of stakeholders?' (chapter 4) and 'Which potential outcome-based quality indicators for patients with COPD can be used for the selection a core set based on the acceptance of stakeholders?' (chapter 5). We first estimated the comparability and discriminability of all outcome measures that could potentially form part of a core set of outcome-based quality indicators. Focus groups were then held with stakeholders, who were invited to accept or decline each measure based on their judgement of its usability and perceived added value as a quality-improvement tool.

In the third step, we aimed to explore the implementation of the set of measurement instruments for patients with COPD in physical therapy practice. The research question was: 'What are the experiences of physical therapists in the use of a standard set for patients with COPD in terms of the interaction between patients, quality-improvement initiatives and public reporting?' (chapter 6). To answer this, we performed a mixed-methods study in two parts: a quantitative survey of 199 physical therapists evaluating the implementation of the standard set, and a qualitative part using individual semi-structured interviews with 11 physical therapists to elucidate their experiences with the use of measurement instruments and the potential use of the data for quality improvement and transparency.

Chapter 2 describes 13 draft outcome domains and associated measures that were rated and discussed by stakeholders in consecutive steps, with a final selection of a core set of five outcome domains and associated measures for patients with NSLBP, and a tool to stratify patients into subgroups. These were: pain measured with a Numeric Pain Rating Scale (NPRS), activities measured with the Patient-Specific Functional Scale (PSFS), physical functioning measured with the Quebec Back Pain Disability Scale (QBPDS) or Oswestry Disability Index (ODI), and perceived treatment effect measured with Global Perceived Effect (GPE-DV). To identify the subgroups, the STarT Back Screening Tool (SBT) was included.

Chapter 3 describes a total of 21 outcome domains and associated measures for patients with COPD that were rated and discussed in the modified RAND/UCLA Delphi procedure, including process measures. Ultimately, eight measures were included in the core set. Four of the measures were mandatory: a process measure characteristic of the practices and physical therapists, the Clinical COPD Questionnaire (CCQ) for quality of life, the Global Perceived Effect (GPE) for experience and the Six-Minute Walk Test (6MWT) for physical capacity. Two of the measures were conditional, to be used when relevant for evaluating specific treatment goals: a hand-held dynamometer (HHD) (with Microfet™) for quadriceps strength, and the Medical Research Council Dyspnea Scale (MRC) for monitoring

dyspnoea. Two exploratory measures were included: accelerometry for physical activity, and the Assessment of Burden of COPD tool (ABC). To identify patient subgroups, a method described in the Dutch standard of care from the Lung Alliance was included.

Chapter 4 presents 15 potential process or outcome quality indicators that were defined for PROMs and associated domains based on the standard set for NSLBP. The comparability and discriminability were described for all potential quality indicators using cohort data. Finally, a core set of quality indicators was selected based on the consensus among stakeholders in focus group meetings. In total, 65,815 completed treatment episodes for patients with NSLBP were provided by 1009 physical therapists from 219 physical therapy practices. The comparability of the data increased after case-mix adjustment, and the discriminability of outcomes between physical therapists or practices was adequate for all 15 potential quality indicators, with intraclass correlation coefficients (ICC) between 0.08 and 0.30. The stakeholders selected a final core set of six quality indicators: two process indicators (the routine measurement of NPRS and the PSFS) and four outcome indicators (pre- and post-treatment score changes for the NPRS, PSFS and QBPDS, and the minimal clinically important difference of the GPE-DV with 95% confidence intervals (CIs)).

Chapter 5 presents two steps. First, a list of potential quality indicators was defined, followed by the determination of the comparability (case-mix adjusted) and discriminability (ICCs) of the quality indicators, analysed using a multi-level analysis. Second, focus group meetings were conducted with physical therapists and senior researchers to select a core set of quality indicators based on their perceived added value as a quality-improvement tool. In total, 229 physical therapists from 137 practices provided 4651 treatment episodes for patients with COPD. The evaluation of the comparability of the quality indicators showed that, in 10 of the 11 case mix–adjusted models, the ICC increased compared with the intercept-only model. An evaluation of the discriminability of the outcomes measured with the quality indicators between physical therapy practices showed that the ICC ranged between 0.01 and 0.34, with five of the 11 ICCs being > 0.10. After discussing the outcomes, the majority of physical therapists in each focus group preferred the inclusion of seven quality indicators in the core set, including three process (proportion of patients with pre- and post-treatment measurements) and three outcome indicators (mean score changes with 95% CI) based upon the 6MWT, the CCQ, and the determination of quadriceps strength using a HHD. Furthermore, a combined process indicator was included to monitor the baseline measurement of three measures used to allocate patients into subgroups.

Chapter 6 describes the results of a sequential explanatory mixed-methods study comprising a survey of 199 physiotherapists and semi-structured interviews with

11 physiotherapists to evaluate their experiences with the implementation of the standard set for patients with COPD. The results showed that, on average, 68.4% of the physical therapists indicated having a positive attitude towards using the standard set of measurement instruments, 85% stated they had sufficient knowledge of the measurement instruments and 84.7% agreed with the items related to contextual factors. The survey demonstrated that 80.3% of the physical therapists thought that the standard set of measurement instruments has an added value for clinical practice and 90.3% indicated that measurement instruments can be of value in evaluating treatment outcomes. During the interviews, the physical therapists mentioned some barriers to the use of the standard set of measurement instruments, such as time and the availability of the standard set in their practice. Physical therapists valued using the measurement instruments for the evaluation of physical therapist treatments for individual patients with COPD, as well as for the aggregated use of outcomes for quality-improvement purposes. Moreover, the physical therapists indicated that the measurement instruments have added value for the anonymized publication of outcomes, providing transparency to policy-makers. Finally, 10 major categories of physical therapist experiences were allocated to seven generic domains of a thematical framework.

Chapter 7 outlines and discusses the results of this thesis by describing the main findings and providing a reflection of the five most essential themes: 1) data collection, 2) development of outcome-based quality indicators, 3) using outcomes in clinical decision-making, 4) using quality indicators for internal quality improvement and 5) using quality indicators for external transparency. Overall, it can be concluded that the elucidation of outcomes using standardised routine data collection of patient-reported outcomes and performance measures can play a key role for quality improvement in physical therapy practice. The standard sets in this thesis were perceived by stakeholders as having added value for the interactions between the patient and physical therapist, for quality-improvement purposes, and for public reporting at the population level as a first step in working towards full transparency. Future efforts should focus on increasing the validity and reliability of the data collected by the participating physical therapists and patients, including patient characteristics that can be used to stratify patients into subgroups. An important prerequisite for the use of outcomes for external transparency at the physical therapist level is to ensure their validity, solve technical issues, and create a system in which patients can provide patient-reported data without interference from their physical therapist. Overall, our conclusion is that the core sets with outcome-based quality indicators with routinely collected aggregated treatment outcomes are able to discriminate between physical therapists or practices. The core sets are accepted by stakeholders and are judged as an important basis for quality improvement and useful in a learning healthcare system.

SAMENVATTING

Hoofdstuk 1

Het routinematig verzamelen van behandeluitkomsten in de eerstelijns fysiotherapie biedt de mogelijkheid om de kwaliteit van deze zorg te monitoren, te evalueren en waar nodig te verbeteren. In een lerend gezondheidszorgsysteem kunnen behandeluitkomsten op verschillende momenten gebruikt worden:

1. In de behandelkamer in de interactie tussen patiënten en fysiotherapeuten voor het stellen van behandeldoelen en het monitoren en evalueren van de behandelresultaten.

2. Op groepsniveau kunnen fysiotherapeuten samen behandeluitkomsten vergelijken en reflecteren op hun eigen behandelresultaten ten opzichte van collega's.

3. Resultaten op groepsniveau kunnen ook worden gebruikt voor externe transparantie, bijvoorbeeld als een hulpmiddel waarmee patiënten een fysiotherapeut kunnen kiezen of voor het inkopen van zorg door zorgverzekeraars.

Voorwaarde voor het gebruik van behandeluitkomsten voor de bovenstaande doelen is dat de gebruikte behandeluitkomsten op een valide en betrouwbare wijze gemeten worden en door de stakeholders (patiënten en patiëntenverenigingen, fysiotherapeuten, beleidsmakers, onderzoekers en zorgverzekeraars) worden gezien als van toegevoegde waarde voor de dagelijkse fysiotherapiepraktijk.

Het doel van dit proefschrift was het ontwikkelen, selecteren en testen van valide en betrouwbare kernsets van kwaliteitsindicatoren in de eerstelijns fysiotherapiepraktijk voor patiënten met aspecifieke lage rugpijn en chronisch obstructieve longziekte (COPD). De kwaliteitsindicatoren in de kernsets omvatten relevante uitkomstdomeinen en bijbehorende meetinstrumenten voor het evalueren van de fysiotherapiebehandeling. De kernsets werden ontwikkeld met minimale datasets voor patiënten met aspecifieke lage rugpijn en COPD. De minimale datasets omvatten routinematig verzamelde behandeluitkomsten uit de dagelijkse eerstelijns fysiotherapiepraktijk. Deze sets zijn op basis van consensus tussen stakeholders vastgesteld.

Naast indicatoren gericht op uitkomsten, zijn er ook structuurindicatoren en procesindicatoren. Structuurindicatoren evalueren de zorgsetting, bijvoorbeeld aanwezigheid van een fitnessruimte in een fysiotherapiepraktijk. Procesindicatoren beschrijven het zorgproces in termen van klinisch redeneren van intake tot afsluiting en het gebruik van meetinstrumenten. In dit proefschrift hebben wij ons met name gefocust op proces- en uitkomstindicatoren gericht op uitkomstdomeinen

gemeten met op PROs of fysieke testen. Hierdoor spreken we van uitkomstgerichte kwaliteitsindicatoren.

De eerste stap in dit promotieonderzoek was het bepalen van uitkomstdomeinen die de geleverde fysiotherapeutische eerstelijns zorg het beste in beeld brengen. Dit kan met behulp van patiënt-gerapporteerde uitkomsten (PROs) of andere patiëntuitkomsten zoals fysieke capaciteit. PROs kunnen worden gemeten met behulp van valide en betrouwbare meetinstrumenten, ook wel patient reported outcome measures (PROMs) genoemd. PROMs zijn vragenlijsten die zich kunnen richten op generieke PROs, bijvoorbeeld pijn, of aandoening specifieke PROs, bijvoorbeeld lichamelijk functioneren bij aspecifieke lage rugpijn. Een voorbeeld van een generiek meetinstrument is de Numeric Pain Rating Scale (NPRS), een voorbeeld van een aandoening specifiek meetinstrument is de Quebec Back Pain Disability Scale (QBPDS). PROMs worden toegepast in de praktijk ter ondersteuning van het patiëntgericht methodisch handelen en het klinisch redeneren, om de uitgangssituatie vast te leggen, behandeldoelen te formuleren en om de voortgang te monitoren. Naast PROMs kunnen ook andere meetinstrumenten, bijvoorbeeld de 6-Minuten WandelTest (6MWT) gebruikt worden voor het meten van fysieke capaciteit bij patiënten met COPD. Verder kunnen patiënten op basis van de resultaten voorkomend uit PROMs en fysieke testen zoals de 6MWT worden onderverdeeld in subgroepen om bij het vergelijken van behandelresultaten rekening te houden met specifieke kenmerken van patiënten.

In **Hoofdstuk 2** beschrijven we de ontwikkeling van een zogeheten minimale dataset voor patiënten met aspecifieke lage rugpijn. Op basis van een literatuurscan hebben we 13 concept uitkomstdomeinen en bijbehorende meetinstrumenten geselecteerd. Met behulp van vragenlijsten, interviews, expertgroep bijeenkomsten en een consensusbijeenkomst hebben patiënten, fysiotherapeuten, beleidsmedewerkers en onderzoekers input gegeven over welke uitkomsten opgenomen moesten worden de minimale dataset. Uiteindelijk werd een minimale dataset van vijf uitkomstdomeinen met bijbehorende meetinstrumenten geselecteerd: pijn gemeten met een NPRS, activiteiten gemeten met de Patient Specifieke Klachten (PSK), fysiek functioneren gemeten met de QBPDS of Oswestry Disability Index (ODI), en het ervaren behandelingseffect gemeten met de Global Perceived Effect (GPE). Daarnaast werd door stakeholders consensus bereikt over het belang van het identificeren van subgroepen op basis van beginmetingen. Omdat de verwachte uitkomsten van de behandeltrajecten voor subgroepen varieert op basis van karakteristieken van de patiënt. Hiervoor werd de STarT Back Screening Tool (SBT) opgenomen in de minimale dataset.

Hoofdstuk 3 beschrijft dezelfde methodiek zoals beschreven in Hoofdstuk 2 om een minimale dataset te ontwikkelen voor patiënten met COPD. Als eerste stap

werd wederom een literatuurscan uitgevoerd waaruit 21 uitkomstdomeinen en bijbehorende meetinstrumenten voor patiënten met COPD zijn geselecteerd. Na dezelfde consensus stappen met stakeholders zoals beschreven in Hoofdstuk 2 werden er uiteindelijk acht uitkomstdomeinen met bijbehorende meetinstrumenten opgenomen in de minimale dataset. Niet alle meetinstrumenten zijn verplicht doordat sommige meetinstrumenten alleen noodzakelijk worden bij bepaalde behandeldoelen. Vier van de meetinstrumenten zijn verplicht: een procesmaat over kenmerken van de fysiotherapeuten en praktijken, kwaliteit van leven gemeten met de Clinical COPD Questionnaire (CCQ), ervaren behandelingseffect gemeten met de Global Perceived Effect (GPE), fysieke capaciteit gemeten met de 6MWT. Twee uitkomstdomeinen met bijbehorende meetinstrumenten zijn voorwaardelijk en worden, indien relevant, gebruikt voor het evalueren van specifieke behandeldoelen: spierkracht van de quadriceps gemeten met een hand-helddynamometer (HHD) (met Microfet™) en kortademigheid gemeten met de Medical Research Council Dyspnea Scale (MRC). Daarnaast zijn twee uitkomstdomeinen met bijbehorende meetinstrumenten opgenomen om te exploreren of deze toegevoegde waarde hebben: fysieke activiteit in het dagelijks leven gemeten met een accelerometer en voor ervaren ziektelast de Assessment of Burden of COPD-tool (ABC-tool). Om subgroepen van patiënten op basis van beginmetingen te identificeren, is een methode opgenomen zoals beschreven in de Nederlandse zorgstandaard van de Long Alliantie.

De meetinstrumenten in de minimale datasets die werden ontwikkeld in hoofdstuk 2 (aspecifieke lage rugpijn) en hoofdstuk 3 (COPD) zijn gebruikt als basis voor het definiëren, testen en selecteren van kwaliteitsindicatoren in hoofdstuk 4 en 5. Een voorbeeld van een kwaliteitsindicator op basis van PROMs is het percentage patiënten met aspecifieke lage rugpijn dat klinisch relevant verbetert op pijn gemeten met de NPRS

In **Hoofdstuk 4** beschrijven we de verschillende stappen die wij doorliepen voor het opstellen van een kernset indicatoren voor het meten van de kwaliteit van de behandeling van patiënten met aspecifieke lage rugpijn. Allereerst werden er 15 mogelijke proces- of uitkomstsindicatoren gedefinieerd, gebaseerd op de uitkomstdomeinen en meetinstrumenten in de minimale dataset. Vervolgens werden deze getest met data uit de dagelijkse praktijk om de vergelijkbaarheid tussen fysiotherapeuten of praktijken te onderzoeken. Hiervoor werden per indicator multilevel regressieanalyses uitgevoerd. Naast vergelijkbaarheid werd ook onderzocht in hoeverre de proces- of uitkomstindicatoren voldoende onderscheidend vermogen hebben tussen fysiotherapeuten of praktijken. In de analyses werd gecorrigeerd voor patiëntkarakteristieken die van invloed zijn op de uitkomsten van de behandeling. Deze zogeheten case-mix correctie bevatte de variabelen geslacht, leeftijd van de patiënt, duur van de klacht voorafgaand aan de eerste zitting en beginscore van

de betreffende uitkomst. De vergelijkbaarheid en het onderscheidend vermogen tussen fysiotherapeuten of praktijken werd berekend met behulp van de intraclass correlatiecoëfficiënten (ICC) voor- en na case-mix correctie.

Aan deze studie deden 1.009 fysiotherapeuten uit 219 fysiotherapiepraktijken mee die 65.815 voltooide behandelepisodes van patiënten met aspecifieke lage rugpijn hebben aangeleverd. Het onderscheidend vermogen van uitkomsten tussen fysiotherapeuten of praktijken was voldoende voor alle 15 mogelijke kwaliteitsindicatoren. De ICC varieerde tussen de 0,08 en 0,30 (0=de variatie in behandeluitkomsten is niet toe te schrijven aan het fysiotherapeutisch handelen, 1= de variatie in behandeluitkomsten is volledig toe te schrijven aan het fysiotherapeutisch handelen). Ten slotte werden fysiotherapeuten en senior onderzoekers uitgenodigd om hun oordeel te geven over de bruikbaarheid en de toegevoegde waarde van de kwaliteitsindicatoren als instrumenten voor kwaliteitsverbetering. In focusgroepen selecteerden fysiotherapeuten en senior onderzoekers een definitieve kernset van zes kwaliteitsindicatoren: twee procesindicatoren (het routinematig meten van NPRS en de PSK), en vier uitkomstindicatoren (verschil tussen het begin en na de behandeling voor de NPRS, PSK en QBPDS, en het minimale klinisch relevante verschil van de GPE-DV met 95% betrouwbaarheidsintervallen (BI)).

In **Hoofdstuk 5** beschrijven we voor patiënten met COPD dezelfde methodiek als beschreven in Hoofdstuk 4. Eerst werd een lijst met mogelijke kwaliteitsindicatoren opgesteld. Vervolgens werd onderzocht of de kwaliteitsindicatoren gebruikt konden worden om praktijken met elkaar te vergelijken en of zij voldoende onderscheid tussen praktijken kunnen aantonen. Om te zorgen dat verschillen ook toegeschreven konden worden aan de praktijken en niet aan patiënteigenschappen werd gecorrigeerd voor leeftijd, geslacht en voor de hoogte van de beginscore van de betreffende uitkomst in de multilevel regressieanalyses.

In totaal hebben 229 fysiotherapeuten uit 137 praktijken data aangeleverd voor de analyses. Het aantal deelnemende fysiotherapeuten ligt lager dan in Hoofdstuk 4 omdat het aantal patiënten met COPD in de praktijken lager was dan het aantal patiënten met aspecifieke lage rugpijn. In totaal werden 4.651 behandelepisodes van patiënten met COPD geanalyseerd. De resultaten werden voorgelegd in focusgroep bijeenkomsten waarin aan fysiotherapeuten en senior onderzoekers werd gevraagd om een kernset van kwaliteitsindicatoren te selecteren op basis van hun toegevoegde waarde als instrumenten voor kwaliteitsverbetering. De evaluatie van de vergelijkbaarheid van de kwaliteitsindicatoren toonde aan dat in 10 van de 11 case-mix-gecorrigeerde modellen de ICC toenam met de correctie in vergelijking met het model zonder case-mix correctie. Het onderscheidend vermogen van de kwaliteitsindicatoren tussen praktijken varieerde, met een ICC van 0,01 tot 0,34. Na bespreking van de uitkomsten gaf de meerderheid van de fysiotherapeuten en

senior onderzoekers in elke focusgroep de voorkeur aan het opnemen van zeven kwaliteitsindicatoren in de kernset. Deze bestaat daarmee uit drie procesindicatoren (het aandeel patiënten met geregistreerde metingen op de 6MWT, de CCQ en de HHD, voor en na de behandeling) en drie uitkomstindicatoren (de gemiddelde verschilscore gebaseerd op de 6MWT voor het meten van fysieke capaciteit, de CCQ voor het meten van kwaliteit van leven, en de HHD voor het meten van de kracht van de quadriceps. Verder werd een procesindicator opgenomen (het aandeel patiënten met beginmetingen op de 6MWT, CCQ en accelerometer) die gebruikt kan worden om patiënten in subgroepen in te delen.

In **Hoofdstuk 6** beschrijven we de resultaten van een onderzoek naar de implementatie van de minimale dataset met meetinstrumenten voor patiënten met COPD in de fysiotherapiepraktijk. Om de ervaringen met het gebruik van de minimale dataset te onderzoeken hebben we een onderzoek uitgevoerd in twee delen. Als eerste hebben we een kwantitatief onderzoek uitgevoerd, met een vragenlijst onder 199 fysiotherapeuten over hun ervaringen met het gebruik van de minimale dataset. Als tweede hebben we een kwalitatief onderzoek uitgevoerd met individuele semigestructureerde interviews met 11 fysiotherapeuten. Het doel van de interviews was om meer inzicht te krijgen in hun ervaringen met het gebruik van de meetinstrumenten en te achterhalen want hun visie was op het mogelijke gebruik van de gegevens voor kwaliteitsverbetering en transparantie.

Uit de resultaten van het kwantitatieve onderzoek bleek dat gemiddeld 68% van de fysiotherapeuten aangaf een positieve houding te hebben ten aanzien van het gebruik van de minimale dataset, 85% gaf aan voldoende kennis van de meetinstrumenten te hebben en 85% was het eens met de items die betrekking hadden op contextuele factoren, zoals het praktijkbeleid en hulp van collega's in de praktijk bij het toepassen van de meetinstrumenten in de minimale dataset. Uit het onderzoek bleek ook dat 80% van de fysiotherapeuten vond dat de minimale dataset een meerwaarde heeft voor de klinische praktijk en 90% geeft aan dat meetinstrumenten van waarde kunnen zijn bij het evalueren van behandelresultaten.

Tijdens de interviews noemden de fysiotherapeuten enkele knelpunten voor het gebruik van de minimale dataset, zoals de benodigde tijd voor het vastleggen en uitvragen van metingen en de beschikbaarheid van de minimale dataset in het elektronische patiëntendossier (EPD) in de praktijk. Fysiotherapeuten waardeerden het gebruik van de meetinstrumenten voor de evaluatie van fysiotherapeutische behandelingen voor individuele patiënten met COPD, evenals voor het geaggregeerde gebruik van uitkomsten voor kwaliteitsverbetering. Bovendien gaven de fysiotherapeuten aan dat de meetinstrumenten een meerwaarde hebben voor het geanonimiseerd publiceren van uitkomsten, waardoor uitkomsten voor patiënten en beleidsmakers transparant worden. Fysiotherapeuten waren sceptisch over

volledige transparantie van uitkomsten voor bijvoorbeeld zorginkoop, met name omdat zij het waarschijnlijk achtten dat andere fysiotherapeuten hun uitkomsten zouden manipuleren om te voorkomen dat hun vergoeding door zorgverzekeraars zou worden aangepast. Deelnemende fysiotherapeuten stelden daarom ook voor om de patiëntmetingen door een derde partij te laten uitvoeren zonder de invloed van fysiotherapeuten.

Vervolgens beschrijven we in **Hoofdstuk 7** de algemene discussie met de belangrijkste bevindingen van dit proefschrift aan de hand van de vijf thema's. Deze thema's zijn: gegevensverzameling, ontwikkeling van uitkomstgerichte kwaliteitsindicatoren, gebruik van uitkomsten in klinische besluitvorming, het gebruiken van kwaliteitsindicatoren voor interne kwaliteitsverbetering en het gebruiken van kwaliteitsindicatoren voor externe transparantie. Over het algemeen kan worden geconcludeerd dat het gebruik van uitkomsten met behulp van gestandaardiseerde routinematige gegevensverzameling van patiëntuitkomsten een sleutelrol kan spelen voor kwaliteitsverbetering in de fysiotherapiepraktijk. De minimale datasets in dit proefschrift werden door stakeholders gezien als van toegevoegde waarde voor de interacties tussen patiënt en fysiotherapeut, voor kwaliteitsverbetering en voor publieke rapportage op populatieniveau als een eerste stap naar volledige transparantie van uitkomsten op het niveau van de fysiotherapeut en/of praktijk.

Toekomstige initiatieven moeten gericht zijn op het vergroten van de validiteit en betrouwbaarheid van de gegevens, inclusief patiëntkenmerken die kunnen worden gebruikt om voor te corrigeren of te stratificeren. Maar moeten initiatieven doorontwikkeld worden waardoor fysiotherapeuten kunnen leren van behandeluitkomsten, zoals een dashboard met feedbackinformatie en ondersteuning bij peer-review bijeenkomsten. Een belangrijke voorwaarde voor het gebruik van uitkomsten voor interne kwaliteitsverbetering en externe transparantie op fysiotherapeutisch niveau is het waarborgen van de validiteit, het oplossen van technische problemen en het creëren van een systeem waarin patiënten uitkomst gerapporteerde gegevens kunnen invullen zonder tussenkomst van hun fysiotherapeut.

Onze conclusie is dat de kernsets met routinematig verzamelde uitkomstgerichte kwaliteitsindicatoren op groepsniveau voldoende onderscheidend vermogen van behandeluitkomsten tussen fysiotherapeuten of praktijken hebben. De kernsets zijn breed gedragen en kunnen in potentie gebruikt worden als belangrijke basis voor kwaliteitsverbetering in een lerend gezondheidszorgsysteem.

DATA MANAGEMENT

All data that is collected in this thesis are stored at the server of the Radboudumc, department IQ-healthcare, secured folder: "H:\PL Philip van der Wees\Project Min dataset Fysio" of which only the project team had access. In this thesis, all patients and physical therapists in the prospective cohort are pseudonymized with a unique code. For the patients, this unique code is obtained by the national registries of the professional bodies, the encrypted keys of the patients are stored in a safe environment of a Third Trusted Party. The unique code of participating physical therapists is the so called "AGB-code", which is a unique identification key for all health professionals in the Netherlands. Personal details (name, date of birth, phone number, email address and unique key) of patients and physical therapists that participated in (focus) groups interviews were stored in a separate secured folder: "H:\PL Philip van der Wees\Project Min dataset fysio NAW" of which only the project leader and PhD candidate had access.

All studies in this thesis were conducted with the principles of Good Clinical Practice, the Netherlands Code of Conduct for Research integrity and according to the Declaration of Helsinki. For each study, we followed the international committee for Research Involving Human Subjects (ICMJE) criteria for authorship. The Medical Ethical Committee of the Radboudumc approved the study protocols of chapter 2 and 3 (registration # 2017-3154) and chapter 4, 5 and 6 (registration # 2019–5455).

Data were analysed in SPSS (quantitative analysis), LimeSurvey version 2.06 (surveys) and Atlas.ti (qualitative analysis). Informed consent was obtained from all individual participants included in the study. According to international standards, data will be stored for 15 years. Data in this thesis are pseudonymized and documented and stored to be reusable after anonymization. After publication of all studies in this thesis data can be reused after a reasonable request upon the PhD candidate.

DANKWOORD

Wie had dat gedacht? Koen die een proefschrift schrijft! In de zomer van 2017 was ik iets heel anders van plan; mijn doel was namelijk om het een periode wat rustiger aan te doen. Ik had de ambitie om mijn werk als fysiotherapeut te combineren met onderzoek doen en kon aan de slag bij IQ-healthcare in het Radboudumc.

Tijdens de kick-off meeting na mijn vakantie bleek ons project groter te zijn dan gedacht. Hierdoor werd het mogelijk om een promotieonderzoeker aan te stellen en al snel kwam de vraag of ik die functie in zou willen vullen. Ik kreeg een week bedenktijd met de mededeling dat als ik het niet ging doen, mijn aanstelling als junior onderzoeker ook helaas zou komen te vervallen. Destijds baalde ik natuurlijk enorm van dit voorstel. Maar achteraf bleek dit precies het duwtje dat ik nodig had om te beginnen aan een prachtig onderzoek in een geweldig team!

Dit proefschrift had ik met geen mogelijkheid kunnen schrijven zonder de hulp, steun en bijdrage van patiënten, fysiotherapeuten, collega's, vrienden en familie. Daar ben ik **iedereen** dan ook **zeer dankbaar** voor. Graag wil ik een aantal mensen in het bijzonder bedanken.

Allereerst mijn promotieteam.

Prof. dr. P.J. van der Wees, beste **Philip**, ik heb je als een zeer betrokken promotor ervaren. Je gaf me vertrouwen en het gevoel dat we dit onderzoek echt samen deden. Ook tijdens de drukke momenten creëerde je structuur en overzicht zodat we als team op de juiste koers bleven. Ik koester de herinnering van onze trip naar de WCPT en onze 'mansion' in centrum van Genève. Bedankt voor de prettige samenwerking waarvan ik hoop dat we die nog lang mogen voortzetten.

Dr. S.A. van Dulmen, beste **Simone**, als dagelijks begeleider vormde je samen met Philip de kern van het promotieteam. Zonder jouw betrokkenheid had ik dit proefschrift niet kunnen schrijven. Jij gaf nooit op, was altijd bereikbaar en bleef op een fijne manier kritisch tot het laatste moment. Daar ben ik je heel dankbaar voor. Zelfs toen we onderweg waren naar een focusgroep in een Twingo zonder airco (met > 35°C) en samen een lekke band moesten verwisselen, waren we nog op tijd!

Dr. H. Kiers, beste **Henri**, wij leerden elkaar kennen in 2013. Je was mijn scriptiebegeleider voor de opleiding fysiotherapie. Toen is mijn interesse in de wetenschap ontstaan. Later zorgde jij ervoor dat ik in contact kwam met Philip, wat uiteindelijk resulteerde in dit promotieonderzoek. Je ervaring, analytische blik en input van 'buitenaf' hebben dit proefschrift absoluut naar een hoger niveau getild. Bedankt voor alle hulp en kansen die jij voor me hebt gecreëerd.

Em. prof. dr. M.W.G Nijhuis-van der Sanden, beste **Ria**, tijdens mijn promotieonderzoek ging je met emeritaat en toch bleef je onderdeel van het team. Je bent betrouwbaar, scherp, kritisch en neemt geen genoegen met half werk. Ondanks een drukke agenda zorgde je er altijd voor dat je goed voorbereid was in onze overleggen. Er is geen versie van het manuscript voorbij gegaan waar jij geen feedback op hebt gegeven. Ik bewonder je arbeidsethos en voel me vereerd dat jij onderdeel was van het promotieteam.

Daarnaast alle **patiënten** en **fysiotherapeuten** die deel hebben genomen aan het project 'minimale dataset lage rugpijn en COPD': voor dit onderzoek was heel veel data nodig, zonder jullie bijdrage was het nooit gelukt, bedankt!

De begeleidingscommissie van het project minimale dataset, **Leen Voogt** (NVVR), **Marie-José Schrasser, Renée Kool** (Long Fonds), **Guus Meerhoff** (KNGF), **Marije de Leur** (SKF), **Rutger Soffers**, **Jan Ypinga** (CZ), **Alida Wolters** en **Sara Meijer** (DFZ). Door jullie bijdrage en (onderlinge) samenwerking hebben we het project succesvol kunnen afronden en verdere implementatie kunnen faciliteren. Onwijs bedankt voor jullie bijdrage.

Mijn collega's van IQ: **Juliette**, **Janine**, **Reinier, Marjo, Anita**, **Angelique** en **Irah,** bedankt voor alle hulp en ondersteuning bij mijn onderzoek. **Pleuntje**, **Joelle**, **Julie**, **Joris**, **Anna**, **Nynke**, **Niek**, **Gijs, Sijmen**, **Amy** en alle andere kelderstrijders, bedankt voor de welkome afleiding, borrels, samenwerking en lunchwandelingen in park Brakkenstein. Überhaupt park Brakkenstein bedankt, je ligt op een perfecte loopafstand waardoor ik veel pauzes van je heb kunnen genieten.

Guus, als mijn scriptiebegeleider hielp je bij mijn eerste vlieguren als onderzoeker. Bedankt voor de eindeloze spreekwoorden, fijne sparsessies en de vriendschap die daaruit voortvloeide.

Thomas, wij zijn pas meer gaan samenwerken tijdens de afronding van mijn proefschrift. Ik bewonder je enthousiasme en energie voor de leerzame onderzoeksprojecten waarin we samenwerken. Ik hoop dat nog lang te mogen doen.

Alle collega's van de **Geeresteingroep**, ondanks dat ik de laatste tijd nog maar beperkt in de praktijk aanwezig ben, voel ik me altijd zeer welkom; dank daarvoor! In het bijzonder **David** en **Jacqueline** voor de ruimte die jullie hebben geboden om mijn onderzoek met patiëntenzorg te kunnen combineren.

Naast het werkende leven zijn vrienden ontzettend belangrijk voor mij om te zorgen voor voldoende joie de vivre! Ik koester dan ook het grote geluk dat ik een ontzettend fijne groep vrienden om mij heen heb. Lieve **Varsity**, **geneco's** (welke geëvalueerd is

tot een **wintersportgroep**), **'Loderlaan boys'**, **Julius**, **Onno**, **Suuz**, **Cecile**, **Britte** en **Joosje**: bedankt voor alle feestjes, verjaardagen, reizen, vakanties, sportactiviteiten, gesprekken en alle andere mooie avonturen.

Lieve **Sybren**, tussen het schrijven door hebben we samen veeeeeeel (koffie) momentjes beleefd waarin ruimte was voor 'ouwe-jongens-krentenbrood', reflectie, ontspanning, sport, onderlinge competitie en humor. Ik ben ontzettend blij met jou als vriend. 'Koooomt goed jong!'

Lieve **Laura** en **Job**, lieve paranimfen, wat zijn jullie fijne mensen en super belangrijk voor mij. Samen met jullie heb ik hoogte- en dieptepunten meegemaakt. Mede door jullie ben ik wie ik ben. 'Iedereen in transitie!'

Lieve **familie**, **Jos** en **Vera**, ik ben blij met jullie, broer en zus. We spreken elkaar niet elke dag, maar ik hou van jullie. **Pa** en **ma**, altijd zijn jullie betrokken en geïnteresseerd in mijn leven. Mede dankzij jullie beschik ik over belangrijke eigenschappen die nodig waren om dit proefschrift tot een succes te maken. Zo hebben jullie me laten zien dat je met discipline, doorzettingsvermogen en samenwerken ver kan komen. Want ik ben een: ….!

En tot slot lieve **Jos**, samen met jou leerde ik om beter te ontspannen en om te genieten van het leven. Je was erbij toen ik in één week een beslissing moest nemen voor dit avontuur. Je steunt me bij alles, onvoorwaardelijk, zo ook tijdens mijn promotieonderzoek. Bedankt voor je liefde, ik hou van jou!

LIST OF PUBLICATIONS

International scientific publications

Guus A. Meerhoff, Arie C. Verburg, Renske M. Schapendonk, Juliette Cruijsberg, Maria W. G. Nijhuis-van der Sanden, Simone A. van Dulmen, Philip J. Van der Wees Reliability, validity and discriminability of patient reported outcomes for non-specific low back pain in a nationwide physical therapy registry: A retrospective observational cohort study *PLoS ONE, 2021,* 10.1371/journal.pone.0251892

Arie C Verburg, Simone A van Dulmen, Henri Kiers, Maria W.G. Nijhuis-van der Sanden, Philip J. van der Wees Patient-Reported Outcome-Based Quality Indicators in Dutch Primary Care Physical Therapy for Patients With Nonspecific Low Back Pain: A Cohort Study *Physical Therapy, 2021, Volume 101, Issue 8 DOI: 10.1093/ptj/pzab118*

Arie C Verburg, Simone A. van Dulmen, Henri Kiers, Jan H.L. Ypinga, Ria MWG Nijhuis-van der Sanden, Philip J van der Wees. Development of a Standard Set of Outcome Domains and Proposed Measures for Chronic Obstructive Pulmonary Disease in Primary Care Physical Therapy Practice in the Netherlands: a Modified RAND/UCLA Appropriateness Method International Journal of Chronic Obstructive Pulmonary Disease 2019:14 2649–2661 DOI: 10.2147/COPD.S219851

Arie C Verburg, Simone A van Dulmen, Henri Kiers, Maria W.G. Nijhuis-van der Sanden, Philip J. van der Wees. Development of a standard set of outcome measures for non-specific low back pain in Dutch primary care physiotherapy practices: a Delphi study *European Spine Journal 2019: volume 28, pages1550–1564 DOI: 10.1007/s00586-019-05962-x*

Other publications

A.C. Verburg, R. Felius, M.W. Heijmans, H. Kiers. Eindrapport resultaatindex fysiotherapie. Zwolle: Stichting Keurmerk Fysiotherapie (SKF), Juni 2022

B. Cijs, A.C. Verburg, N.M. Swart, C. Veenhof, P.J. van der Wees. KNGF-standpunt Fysiotherapie bij COVID-19, aanbevelingen voor fysiotherapeutisch handelen in de eerste lijn, versie 3.0, 1 maart 2022. Amersfoort: Koninklijk Nederlands Genootschap voor Fysiotherapie (KNGF); 2022.

Verburg A.C., Dulmen van S.A., Wees van der P.J. Ontwikkeling en gebruik van uitkomstindicatoren in de fysiotherapie. *Physios, december 2021, nummer 4*

Mitchell van Doormaal, Danielle Conijn, Koen Verburg, Jesper Knoop, Philip van der Wees. Deel 1: aandoeningoverstijgende richtlijnen; Niet de aandoening maar het fysiek functioneren centraal, FysioPraxis November 2021

Koen Verburg, Simone van Dulmen, Henri Kiers, Ria Nijhuis-van der Sanden, Philip van der Wees. Development of a standard set of outcome domains and proposed measures for chronic obstructive pulmonary disease in primary care physical therapy practice in the Netherlands. FysioPraxis November 2020

Dulmen van S.A., Verburg A.C., Cruijsberg J., Wees van der P.J. Eindrapport toepassing van minimale dataset en kwaliteitsindicatoren voor lage rugklachten in de fysiotherapie. Nijmegen: IQ healthcare, april 2020

Verburg AC, Dulmen van SA, Cruijsberg J, Wees van der PJ. Eindrapport toepassing van minimale dataset en kwaliteitsindicatoren voor patiënten met COPD in de fysiotherapie. Nijmegen: IQ healthcare, juni 2020

van Dulmen S.A., van der Wees P.J., Verburg A.C., Nijhuis-van der Sanden M.W.G. Ontwerp van een minimale dataset voor lage rugklachten en COPD in de fysiotherapie. Nijmegen, IQ healthcare, 2017

Verburg A.C., Dulmen van S.A., Wees van der P.J. Ontwikkeling en gebruik van uitkomstindicatoren in de fysiotherapie. *Physios, 2021, nummer 4*

Awards and grants

BackCare Award BackCare Award, chosen by the Executive Committee from the top 10 scoring papers submitted the previous year - Society of Back Pain Research jan. 2019

New Investigator Award New Investigator Award, Society of Back Pain Research dec. 2018

Research profiles

https://www.researchgate.net/profile/Ac-Verburg
https://www.linkedin.com/in/koen-verburg-39544a69

ABOUT THE AUTHOR

Koen Verburg was born on April 7th, 1991 in Amersfoort, the Netherlands. He completed secondary school in 2009 at the Koningin Wilhelmina College in Culemborg. In 2013, he graduated his study physical therapy at the HU University of Applied Sciences in Utrecht (BSc), followed by Clinical Health Sciences in 2017, focus: Physical Therapy Science, at Utrecht University (MSc.)

His professional career started in 2013 as physical therapist at Rijndam rehabilitation center in Rotterdam (stopped in 2014), and primary care practice Geeresteingroep in Woudenberg. At the Geeresteingroep he has treated various patient populations and was physical therapist of multiple sport teams (speed skating, field hockey, tennis and korfball). Currently, his physical therapy treatment at the Geeresteingroep is focused on patients with hand complaints in collaboration with certified hand therapists in Soest and with (plastic) surgery departments of Meander Medical Center in Amersfoort.

For his master he started an internship at IQ healthcare in 2016 which resulted in a position as junior researcher, followed by a PhD project in 2017 that resulted in this thesis. Besides the PhD, he is currently involved in several research projects in allied healthcare, such as the evaluation of allied healthcare in Dutch patients recovering from COVID-19, development of clinical guidelines, and projects focusing on development of tangible methods to learn from routinely collected outcomes in daily practice.

In 2021 he started (part-time) working as a policy advisor at the Association for Quality in Physical Therapy (SKF). His projects are strongly related to the topic of his PhD thesis. He is working together with patient (organisations), physical therapists, the Royal Dutch Society for Physiotherapy (KNGF), software companies and health insurers concerning the implementation, evaluation and innovation of routinely collected outcomes in daily practice.

PHD PORTFOLIO

Institute for Health Sciences
Radboudumc

Name PhD candidate: A.C. (Koen) Verburg	**PhD period:** 1-10-2017 to 01-02-2022
Department: Scientific Centre for Quality of Healthcare	**Promotor:** Prof. dr. P.J. (Philip) van der Wees Prof. dr. M.W.G. (Ria) Nijhuis – van der Sanden
Graduate School: Radboud Institute for Health Sciences	**Co-promotors:** Dr. S.A. (Simone) van Dulmen Dr. H. (Henri) Kiers

	Year(s)	ECTS
TRAINING ACTIVITIES		
a) Courses & Workshops		
• Radboudumc introduction, Nijmegen, the Netherlands	2017	0.5
• RIHS introductory course, Radboudumc, Nijmegen, the Netherlands	2017	0.75
• PhD retreat, RIHS, 's Hertogenbosch, the Netherlands	2018	1.0
• Statistics for PhD candidates, Radboud University, Nijmegen, the Netherlands	2018	2.0
• Journal club klinimetrie, IQ Healthcare, Radboudumc, Nijmegen, the Netherlands	2019	2.0
• Scientific Integrity course, Radboudumc, Nijmegen, the Netherlands	2019	1.0
• Journal club COSMIN checklist, IQ Healthcare, Radboudumc, Nijmegen, the Netherlands	2019	2.0
• Kwalitatieve onderzoeksmethoden, IQ Healthcare, Radboudumc, Nijmegen, the Netherlands	2019	1.0
• PhD intervisie IQ Healthcare, Radboudumc, Nijmegen, the Netherlands	2018-2020	3.0
• Academic writing for PhD candidates, Radboud University, Nijmegen, the Netherlands	2020	1.0
• Mixed model analysis EpidM, VU University Medical Center	2019	2.0
• Workshop multilevel analyses, IQ Healthcare, Radboudumc, Nijmegen, the Netherlands	2019	0.4

• Introduction course qualitative research, IQ Healthcare, Radboudumc, Nijmegen, the Netherlands	2019	2.0
• Basiscursus Regelgeving en Organisatie voor Klinisch Onderzoekers (BROK-cursus), Radboudumc, Nijmegen, the Netherlands	2020	1.5
b) Seminars & lectures		
• Introduction course participants MDS project (online)	2018	0.1
• Lecture for physical therapy network for patients with COPD	2019	0.1
c) Symposia & congresses		
Oral presentations		
• Dutch Physical Therapy day congress (KNGF), Brabanthallen, 's Hertogenbosch, the Netherlands	2022	1.0
• Society for Back Pain Research, Groningen, the Netherlands	2018	1.0
Poster presentations		
• Dutch Physical Therapy day congress (KNGF), Brabanthallen, 's Hertogenbosch, the Netherlands	2019	0.1
• Society for Back Pain Research, Sheffield, England	2019	1.0
• World Congress Physical Therapy, Geneva, Switzerland	2019	3.0
• Participation	2018	0.3
• WCF (Scientific College for Physiotherapy) day, Amersfoort, the Netherlands WCF (Scientific College for Physiotherapy) day, Hilversum, the Netherlands	2019	0.3
• WCF (Scientific College for Physiotherapy) day, Deventer, the Netherlands	2021	0.3
Organizing		
• Course for participants of research project MDS	2019-2020	5.0
TEACHING ACTIVITIES		
Lecturing		
• Introduction in qualitative research (master medicine students)	2019-2020	3.0
Supervision of internships		
• Supervision UU students (master thesis Clinical Health Sciences, direction Physical Therapy Science)	2018-2021	8.0
• Supervision RU students (master Biomedical Sciences)	2019-2020	4.0
• Internship 'Meet your PhD' (bachelor Biomedical Sciences)	2018-2019	0.4
	TOTAL	**46.75**

Radboud University Radboudumc university medical center

Institute for Health Sciences

Radboudumc