Cost of Illness due to Back Pain in Older People:

Healthcare utilization, modifiable prognostic factors, and the measurement properties of self-reported productivity loss using the iMTA Productivity Cost Questionnaire

Rikke Munk Killingmo

OsloMet Avhandling 2022 nr. 36

OSLO METROPOLITAN UNIVERSITY
STORBYUNIVERSITETET

Cost of Illness due to Back Pain in Older People

Healthcare utilization, modifiable prognostic factors, and the measurement properties of self-reported productivity loss using the iMTA Productivity Cost Questionnaire

Rikke Munk Killingmo

THESIS FOR THE DEGREE OF PHILOSOPHIA DOCTOR
Department of Rehabilitation Science and Health Technology
Faculty of Health Science
OsloMet - Oslo Metropolitan University

Autumn 2022

OsloMet Avhandling 2022 nr 36

ISSN 2535-471X
ISBN 978-82-8364-433-3

OsloMet – storbyuniversitetet
Universitetsbiblioteket
Skriftserien
St. Olavs plass 4,
0130 Oslo,
Telefon (47) 64 84 90 00

Postadresse:
Postboks 4, St. Olavs plass
0130 Oslo

Trykket hos Byråservice

Trykket på Scandia 2000 white, 80 gram på materiesider/200 gram på coveret

Livet forstås baglæns, men må leves forlæns

Søren Kierkegaard

Big problems are best solved in small pieces

Frank Sonnenberg

ACKNOWLEDGEMENTS

This thesis is the result of the hard work and creative input of many people to whom I am greatly indebted. First and foremost, I would like to thank my supervisors, Margreth Grotle and Kjersti Storheim.

Margreth, the list of things for which I'm grateful to you is long. Thank you for introducing me to your world of research, and for many hours of guidance, support, and inspiration over the years. Thank you for believing in me, for pushing me through constructive feedback, for raising the bar, and for all the work you've done behind the scenes to create opportunities for me. You are a warm and caring person, your door is always open regardless of how busy you are, and your passion for research is infectious. It has been a privilege to have had you as my main supervisor, and I hope to be able to work with you in the years to come.

Kjersti, thank you for your support, for sharing your knowledge, and for providing essential guidance on all aspects of my work. I highly value your expertise in research and your impressive overview of the field. It has been a privilege to have had you as my co-supervisor.

I would also like to thank the international BACE consortium and the scientific board of BACE-N: Margreth Grotle, Chris Maher, Bart W. Koes, Manuela Ferriera, Kjersti Storheim, Liv Inger Strand, Silje Stensrud, Iben Axen and Thor Einar Holmgaard. I am deeply grateful to all of you for your invaluable contributions in planning BACE-N, which have made this thesis possible.

To my co-authors, thank you for your hard work and creative input: Anne Therese Tveter, Danielle van der Windt, Sita M. A. Bierma-Zeinstra, and Zinajda Zolic-Karlssson. Special thanks to Ørjan N. Vigdal and Lise Kretz who have been my "partners in crime" in the BACE-N project, to Bart W. Koes and Alessandro Chiarotto for giving me the opportunity to have a memorable research stay at the Erasmus University Medical Center in the Netherlands, and to Milada C. Småstuen, for sharing your expertise on statistics, a brief visit to your office has kept me on track many times over the years.

To the many people who assisted in recruitment and data collection for the BACE-N project, my sincere thanks for doing an incredible job despite little or no reward, without you the BACE-N project would have been impossible to complete: Jan Harald Lønn, Kerstin Ulrich, Christoffer Børsheim, Lise Lothe, Tim Raven, Alexander Tingulstad, Solveig Flugstad, Lise Grete Kjønø, Emil Thomassen, Øyvind Høvik, Andreas Hoff Nordvik, Anna Allen-Unhammer, Morten Nilsen, Philip Wilkens, Kathrine N. Mikkelsen, Joakim Hære, Ingrid Misje, Ingeborg Landstad and Unicare Friskvern, Are Hansen, Aleksander M. Killingmo, Kristine Ø. Tollefsen, Trond Magne Aasberg, Simon Vandås, Haakon Asmyhr, Daniel Ekeberg, Renate Meier, Mona Øversveen, Stina Lund, Marte A. Paulsen, Andrea Kolstad, Berte Marie Enger, Marie Torgstad, Palwinder Singh, Astrid Figger, Bjørn Tore Bjørkedal, Lasse Christensen, Ambroise Omgba, Christian Mayer, Ane Klevberg, Kaja McCormick, Morten de la Cruz, Geir Haram, Bent Ulseth, Jorunn Salater, Svein Erik Sandelien, Harald Nordheim, Geir Erik Johansen, Agnes C. Mordt and her colleagues at Stadion Fysikalske Institutt og Treningssenter, Ola Sand, Rebecka Jensen, Anita Tveiten, Petter Narvesen, Torstein B. Løksa, Ragnhild Perstølen, Steinar Forshei, Mette Brekke, Daniel Major, Eli Magnesen, Johan Edvard Tellum, Joakim Ordahl, Jørgen Kongtorp, Bård Kvam, Olav Aase, Lars Martin Fischer, Martin Haagensen, Nanna B. Rykken, Morten Veening, Thea Tømmervåg, Jørgen Øyen, Håvard Nordås, Tore Viste-Ollestad, Bjørn Røe, Jack Johnson, Carl-Erik Høgquist, Ingrid Hystad, Mats Thorbeck, Elisabeth Barø, Vaiva Grube, Sondre Stangeland, Cathrine Rossland, Svein Aarseth, Alexander Diesen, Marianne Storberget,

Mathias Svanevik, Lars Gullestad, Ida Svalstuen, Nora Helk, Solveig Hjertaas, Lene Faannessen, Grete Bråten, Annecken L. Haugen, Hege Herstad, Cathrine Natland, Frøydis B. Asbø, Ola Klaastad, Thorleif Henning Monsen, Geir Wiik, Tonje H. Mysen, Jørn Christian Halvorsen, Tine Tandberg, Eir Marie Bergan, Rune Solheim, and Ole Kristoffer Larsen.

To the MI-NAV project at Oslo Metropolitan University, thank you for including one of my research questions in your project and providing data for Paper IV.

I would also like to express my gratitude to all the study participants. Thank you for your time and effort, without you there would be no projects.

To everyone at the MUSK Health research group at Oslo Metropolitan University, thank you for including me in an amazing and friendly research environment. Thank you for all the help, inspiration, and good times. Special thanks to Margreth Grotle and Britt Elin Øiestad for great leadership, and for creating a positive and stimulating research culture. You make MUSK Health a great place to be. And thank you to the "vol. 1 group of MUSK Health PhD students" for valuable times and friendship: Ørjan N. Vigdal, Daniel Major, Henriette Jahre, Kaja Smedbråten, Tarjei Rysstad, Fiona Aanensen, Alexander Tingulstad, and Mette Kaarbø.

To my family and friends, I am sincerely grateful for your interest, support, and patience. To my mom, thank you for your encouragement and for introducing the attitude "you can accomplish anything if you are willing to work long and hard enough". To my dad, thank you for being a reflective farmer from Jutland and for quiet discussions on common sense. To my beloved sister, thank you for your unconditional support and for always being proud of me.

Last, but certainly not least, I'd like to thank my dear husband, Aleksander, and my two children, Vilde and Kasper. Aleksander, my favourite person in the world, words cannot describe how grateful I am for your love, support, and caring. Thank you for being my safe haven and for reminding me to enjoy life. Vilde and Kasper, thank you for always reminding me of the most important side of life. You mean the world to me!

Thank you all!

If I have seen further than others,
it is by standing upon the shoulders of giants

Isaac Newton

Rikke Munk Killigmo

Oslo, May 2022

The work on this thesis was performed at the Oslo Metropolitan University in the period from 2016-2022. The main funding source of this thesis was the Oslo Metropolitan University. Parts of data collection for Papers I-IV were additionally funded by the Norwegian Fund for Post-Graduate Training in Physiotherapy, A life in movement - the Norwegian chiropractors' research foundation, the Research Council of Norway, the department of General Practice at Erasmus Medical Center, Rotterdam, the Netherlands, and the Coolsingel foundation, Rotterdam, the Netherlands.

SUMMARY

Introduction Back pain imposes substantial individual and societal costs, and with an aging population and an increasing number of older people with back pain, these costs are likely to increase in the years to come**.** To improve the use of scarce healthcare resources and reduce the economic burden on our healthcare systems, it is vital to map healthcare utilization and related costs and identify modifiable prognostic factors of the high costs related to healthcare utilization. No such studies have been conducted among a sample of exclusively older people, although the prevalence of seeking healthcare for back pain increases with age. Furthermore, to promote comprehensive healthcare economic evaluations, there is a need for valid generic instruments for measuring productivity costs. Productivity costs often reflect a large part of total costs related to health and healthcare interventions.

Aims The primary aim of this thesis was to develop new knowledge on the cost of illness due to back pain among older people, to describe healthcare utilization and estimate associated costs among older people seeking primary care due to back pain (Paper I) and to identify modifiable prognostic factors of high costs related to healthcare utilization (Paper II). A secondary aim was to evaluate the measurement properties of the iMTA Productivity Cost Questionnaire (iPCQ) (Papers III and IV).

Methods Papers I and II were conducted using a cohort study design with one-year of follow-up. Data from the Back Complaints in the Elders (BACE) consortium in Norway (BACE-N) and the Netherlands (BACE-D) were used. BACE-N included 452 people aged ≥ 55 years seeking Norwegian primary care with a new episode of back pain, and BACE-D included 675 people aged > 55 years seeking Dutch primary care with a new episode of back pain. In Paper I, healthcare utilization and related costs were described for the whole BACE-N sample as well as for patients with different risk profiles according to the STarT Back Screening Tool (SBST). In Paper II, potential modifiable prognostic factors of high costs related to healthcare utilization were identified in BACE-N, and the findings were then replicated in BACE-D. In Papers III and IV, the content validity, construct validity, criterion validity, and test-retest reliability of the iPCQ were evaluated in two different Norwegian samples. Paper III was a cross-sectional study, including a test-retest assessment, of 115 patients with musculoskeletal disorders referred to an outpatient rehabilitation clinic. Paper IV was a cross-sectional study, including one year of retrospective public registry data on absenteeism, among 144 people who had been on sick leave for at least four weeks due to musculoskeletal disorders.

Main results and conclusions In Paper I, the one-year mean and median total cost per patient were estimated at €825 and €364, respectively. The largest cost category was primary care consultations (56% of total costs). Imaging rate was 34%. A total of 34-45% of patients used medication, and the most widely used type was paracetamol (27-35% of patients). Patients with medium and high risk of persistent disabling back pain according to the SBST had a significantly higher degree of healthcare utilization compared to patients with low risk. In Paper II, four modifiable prognostic factors associated with high costs related to healthcare utilization were identified and replicated: a higher degree of pain severity, disability, and depression, and a lower degree of physical health-related quality of life. In Paper III, the content validity of the iPCQ was found to be sufficient, construct validity was confirmed, and test-retest reliability was acceptable. In Paper IV, self-reported productivity loss by the iPCQ showed good agreement with public registry data regarding the occurrence and duration of long-term absenteeism. However, the iPCQ does not cover part-time sick leave and overestimated the number of days with complete absenteeism with median 17 days.

Implications This thesis provides knowledge that can be used to inform the use of scarce healthcare resources and reduce the economic burden of back pain on healthcare systems. Decreasing the use of imaging and paracetamol seem to be important areas for quality improvement in primary care management of older patients with back pain. In addition, pain severity, disability, depression, and physical health-related quality of life are potential target areas for interventions directed towards reducing high costs related to healthcare utilization among these patients. Moreover, this thesis contributes knowledge useful for conducting comprehensive health economic evaluations which include productivity costs. The iPCQ can be recommended as a useful tool for measuring three important components of productivity costs among people with musculoskeletal disorders: absenteeism, presenteeism, and costs related to unpaid work.

SAMMENDRAG

Bakgrunn Ryggsmerter medfører betydelige individuelle og samfunnsmessige kostnader. Med utgangpunkt i en aldrende befolkning og et økende antall eldre med ryggsmerter, forventes disse kostnadene å øke i årene som kommer. For å forbedre bruken av begrensede helseressurser og redusere den økonomiske belastningen på helsevesenet, er det viktig å kartlegge helsetjenesteforbruk, relaterte kostnader samt å identifisere modifiserbare prognostiske faktorer for høye kostnader relatert til helsetjenesteforbruk. Ingen slike studier har blitt utført på et utvalg av utelukkende eldre personer, til tross for at forekomsten av å søke helsehjelp for ryggsmerter øker med alderen. For å fremme helhetlige helseøkonomiske vurderinger, er det behov for valide generiske instrumenter som måler produktivitetskostnader. Produktivitetskostnader reflekterer ofte en stor del av de totale kostnader knyttet til helse- og helsetiltak.

Formål Det overordnede formålet med denne avhandlingen var å utvikle ny kunnskap om kostnader relatert til ryggsmerter blant eldre personer; å beskrive helsetjenesteforbruk og estimere relaterte kostnader blant eldre personer som søker primærhelsetjeneste på grunn av ryggsmerter (artikkel I) samt å identifisere modifiserbare prognostiske faktorer for høye kostnader relatert til helsetjenesteforbruk (artikkel II). Et sekundært formål var å evaluere måleegenskapene til «the iMTA Productivity Cost Questionnaire (iPCQ)» (artiklene III og IV).

Metode Artiklene I og II ble utført ved bruk av et kohortstudiedesign med ett års oppfølging. Data fra «Back Complaints in the Elders (BACE)» konsortiet i Norge (BACE-N) og Nederland (BACE-D) ble brukt. BACE-N inkluderte 452 personer i alderen ≥ 55 år som søkte norsk primærhelsetjeneste på grunn av en ny episode med ryggsmerter. BACE-D inkluderte 675 personer i alderen > 55 år som søkte nederlandsk primærhelsetjeneste på grunn av en ny episode med ryggsmerter. I artikkel I ble helsetjenesteforbruk og relaterte kostnader beskrevet for hele BACE-N utvalget og for personer med ulike risikoprofiler i henhold til «the STarT Back Screening Tool (SBST)». I artikkel II ble potensielle modifiserbare prognostiske faktorer for høye kostnader relatert til helsetjenesteforbruk identifisert i BACE-N, deretter ble funnene replikert i BACE-D. I artiklene III og IV ble innholdsvaliditet, konstruktvaliditet, og kriterievaliditet samt test-retest reliabilitet av iPCQ evaluert i to forskjellige norske utvalg. Artikkel III var en tverrsnittstudie, med en test-retest-vurdering, av 115 pasienter med muskel- og skjelettplager henvist til poliklinisk rehabiliteringsklinikk. Artikkel IV var en

tverrsnittstudie, inkludert offentlig registerdata for sykefravær målt ett år retrospektivt, av 144 personer som hadde vært sykemeldt i minst fire uker på grunn av muskel- og skjelettplager.

Hovedresultater og konklusjoner I artikkel I ble ett års gjennomsnitts- og median totalkostnad per pasient estimert til henholdsvis €825 og €364. Den største kostnadskategorien var primærhelsetjenestekonsultasjoner (56% av totale kostnader). Bildediagnostikk frekvensen var 34 %. Totalt 34-45% av pasientene brukte medisiner, den mest brukte typen var paracetamol (27-35 % av pasientene). Pasienter med middels og høy risiko for vedvarende funksjonsbegrensende ryggsmerter, ifølge SBST, hadde et signifikant høyere helsetjenesteforbruk sammenlignet med pasienter med lav risiko. I artikkel II ble fire modifiserbare prognostiske faktorer assosiert med høye kostnader relatert til helsetjenesteforbruk identifisert og replikert: en høyere grad av smerteintensitet, funksjonsbegrensninger og depresjon samt en lavere grad av fysisk helserelatert livskvalitet. I artikkel III ble innholdsvaliditeten til iPCQ funnet å være tilstrekkelig, konstruktvaliditeten ble bekreftet, og test-re-test reliabiliteten var akseptabelt. I artikkel IV viste selvrapportert produktivitetstap målt med iPCQ godt samsvar med offentlig registerdata for forekomst og varighet av langtidssykefravær. Imidlertid dekker ikke iPCQ deltidssykefravær og overestimerte antall dager med fullstendig sykefravær med median 17 dager.

Implikasjoner Denne avhandlingen gir kunnskap som kan brukes til å informere bruken av knappe helseressurser og redusere den økonomiske belastningen av ryggsmerter på helsevesenet. Å redusere bruken av bildediagnostikk og paracetamol ser ut til å være viktige områder for kvalitetsforbedring i primærhelsetjenestebehandlingen av eldre pasienter med ryggsmerter. I tillegg er smerteintensitet, funksjonsbegrensninger, depresjon og fysisk helserelatert livskvalitet mulige målområder for intervensjoner rettet mot å redusere høye kostnader knyttet til helsetjenesteforbruk blant disse pasienter. Videre gir denne avhandlingen kunnskap som kan fremme gjennomføringen av helhetlige helseøkonomiske vurderinger, hvor produktivitetskostnader er inkludert. Instrumentet iPCQ kan anbefales som et nyttig verktøy for å måle tre viktige komponenter av produktivitetskostnader blant personer med muskel- og skjelettplager: sykefravær, sykenærvær og kostnader knyttet til ulønnet arbeid.

TABLE OF CONTENTS

LIST OF PAPERS

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I. Killingmo RM, Storheim K, van der Windt DA, Zolic-Karlsson Z, Vigdal ØN, Kretz L, Småstuen MC, Grotle M. *Healthcare utilization and related costs among older people seeking primary care due to back pain: findings from the BACE-N cohort study*. BMJ Open 2022 Jun 20;12(6):e057778. DOI: 10.1136/bmjopen-2021-057778

II. Killingmo RM, Chiarotto A, van der Windt D, Storheim K, Bierma-Zeinstra SMA, Småstuen MC, Zolic-Karlsson Z, Vigdal ØN, Koes BW, Grotle M. *Modifiable prognostic factors of high costs related to healthcare utilization among older people seeking primary care due to back pain: an identification and replication study*. BMC Health Serv Res. 2022 Jun 18;22(1):793 DOI: 10.1186/s12913-022-08180-2

III. Munk R, Storheim K, Småstuen MC, Grotle M. *Measuring Productivity Costs in Patients With Musculoskeletal Disorders: Measurement Properties of the Institute for Medical Technology Assessment Productivity Cost Questionnaire*. Value Health. 2019; 22(12):1410-1416. DOI: 10.1016/j.jval.2019.07.011

IV. Killingmo RM, Tveter AT, Småstuen MC, Storheim K, Grotle M. *Comparison of self-reported and public registered absenteeism among people on long-term sick leave due to musculoskeletal disorders: criterion validity of the iMTA Productivity Cost Questionnaire*. Eur J Health Econ. 2021 Aug; 22(6):865-872. DOI: 10.1007/s10198-021-01294-0

ABBREVIATIONS

Table A1. Abbreviations

BACE	The Back Complaints in the Elders (international consortium)	
BACE-D	The Back Complaints in the Elders study - the Netherlands	
BACE-N	The Back Complaints in the Elderly - Norway study	
BCa	Bias-corrected and accelerated	
BP	Back pain	
CES-D	The Center for Epidemiological Studies-Depression questionnaire	
CI	Confidence interval	
CLBP	Chronic low back pain	
COI	Cost of illness	
COSMIN	The COnsensus-based Standards for the selection of health Measurement INstruments	
CT	Computerized tomography	
DALY	Disability adjusted life years	
GBD	Global Burden of Disease	
GP	General practitioner	
EPV	Events per variable	
FABQ-PA	The Fear Avoidance Beliefs Questionnaire - Physical Activity subscale	
ICC	Intraclass correlations coefficient	
ICPC code	International Classification of Primary Care code	
iMTA	The Institute for Medical Technology Assessment	
iPCQ	iMTA Productivity Cost Questionnaire	
iPCQ-VR	iMTA Productivity Cost Questionnaire - Vocation rehabilitation	
IQR	Interquartile range (25^{th} percentile - 75^{th} percentile)	
LBP	Low back pain	
MI-NAV study	A randomized controlled trial of the effectiveness of adding motivational interviewing or stratified vocational advice intervention to usual case management on return to work for people with musculoskeletal disorders in Norway	
MRI	Magnetic resonance image	
NA	Not applicable	
NAV	The Norwegian Labour and Welfare Administration	
NoMA	Norwegian Medicines Agency	
NRS	Numeric Rating Scale	
NSAID	Non-steroidal anti-inflammatory drugs	
NSD	The Norwegian Social Science Data Service	
OR	Odds Ratio	
PROGRESS	The PROGnosis RESearch Strategy	
PWQ	The Physical Workload Questionnaire	
QOL	Quality of life	
QPS Nordic	General Nordic questionnaire for psychological and social factors at work	
REC	The Norwegian Regional Committee for Medical Research Ethics	
REMARK	REporting recommendations for tumour MARKer prognostic studies	
RMDQ	The Roland Morris Disability Questionnaire	
SBST	The Keele STarT Back Screening Tool	
SCQ	The Self-Administered Comorbidity Questionnaire	
SD	Standard deviation	
SF-36	36-Item Short Form Health Survey	
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology	
YLD	Years lived with disability	
Currencies used within thesis		
€	Euro	
£	British pound	Exchange rate February 2020, €1 = £0.9
$ / USD	United States dollar	Exchange rate February 2020, €1 = $1
AU$	Australian dollar	Exchange rate February 2020, €1 = AU$ 1.7
CAD	Canadian dollar	Exchange rate February 2020, €1 = CAD 1.4
NOK	Norwegian krone	Exchange rate February 2020, €1 = NOK 10
SEK	Swedish krona	Exchange rate February 2020, €1 = SEK 11
TL	Turkish lira	Exchange rate February 2020, €1 = TL 7
Yen	Japanese Yen	Exchange rate February 2020, €1 = Yen 120

DEFINITIONS AND MAIN CONCEPTS

Table A2. Definitions and main concepts

Back pain	A symptom defined by the location of pain, typically between the lower rib margins and the inferior gluteal folds [1]
Bias	A difference (deviation) in a particular direction (systematic) between the results of a study (parameter estimate) and what happens in real life (population value) [2]
Burden of disease	The total, cumulative consequences of a defined disease with respect to disability and premature mortality [3]
Construct validity	The degree to which the scores of a measurement instrument are consistent with hypotheses based on the assumption that the measurement instrument validly measures the construct to be measured. Aspects of measurement property: *Structural validity:* the degree to which the scores of a measurement instrument are an adequate reflection of the dimensionality of the construct to be measured. *Hypotheses testing:* the degree to which the scores of a measurement instrument are consistent with hypotheses based on the assumption that the measurement instrument validly measures the construct to be measured. *Cross-cultural validity:* The degree to which the performance of the items on a translated or culturally adapted measurement instrument are an adequate reflection of the performance of the items of the original version of the measurement instrument [4]
Content validity	The degree to which the content of a measurement instrument is an adequate reflection of the construct to be measured. Aspect of measurement property: *Relevance:* Are the included items relevant for the construct of interest, the target population of interest, and the context of use of interest? Are the response options appropriate? Is the recall period appropriate? *Comprehensiveness:* Are all key concepts included? *Comprehensibility:* Are the instruction, items, and response options understood by the population of interest as intended? Are the items appropriately worded? Do the response options match the questions? [4, 5]
Cost of illness studies	A descriptive study that identifies and estimate costs of a defined illness or health problem [6] Main concepts and approaches used: *Cost categories: Direct costs:* monetary costs related to healthcare (as diagnosis, treatment, rehabilitation, etc.) and non-healthcare costs related to the provision of healthcare (as transportation). *Indirect costs:* monetary costs related to productivity loss resulting from illness or treatment, typically divided into three domains: absence from paid work (absenteeism), reduced productivity while at paid work (presenteeism), and productivity costs related to unpaid work (such as household work, care work, and volunteer work). *Intangible costs:* related to discomfort, pain, anxiety, or inconvenience, usually measured by reduction in quality of life [6-10]. *Cost perspectives:* The cost perspective indicates who bears the costs, thus determine which costs are to be included. A *societal perspective* includes all costs, regardless of who incurs them. A *health system perspective* includes costs borne by the health sector. An *employer or insurer perspective* includes only costs paid for by that organization [6, 7]. *Costing direction:* Cost of illness studies are performed either "top down" or "bottom up". *Top-down studies:* statistical databases and registries are used to estimate costs for a given prevalence sample, providing aggregate data at a national level. Limitations with a top-down approach: some costs are not available from these source and total costs will be underestimated, databases may be incomplete, or items miscoded. *Bottom-up studies:* costs are collected directly from a patient sample; results can be extrapolated using prevalence data to estimate costs at a national level. Limitations with a bottom-up approach: ensuring an unbiased and representative sample of the overall patient population, recall bias [6].
Covariate	A variable that explains a part of the variability in the outcome [11]
Criterion validity	The degree to which the scores of a measurement instrument are an adequate reflection of a gold standard [4]
Internal consistency	The degree of the interrelatedness among the items [4]
Prognostic factor	Any measure that, among people with a given health condition (that is, a startpoint), is associated with a subsequent clinical outcome (an endpoint) [12] The PROGRESS (PROGnosis RESearch Strategy) framework classifies prognosis research into four main types of study: 1) overall prognosis research, 2) prognostic factor research, 3) prognostic model research, and 4) predictors of treatment effect research [12-15]
Reliability	The extent to which a measurement instrument for patients who have not changed are the same for repeated measurement under several conditions [4]
Validity	The degree to which a measurement instrument measures the construct(s) it purports to measure [4]

INTRODUCTION

Back pain is the leading cause of disability globally [16, 17], and alongside a high disease burden, back pain also creates a major socio-economic burden related to healthcare utilization and productivity loss [18-20]. Back pain imposes substantial individual and societal costs, and with an aging population and an increasing number of older people with back pain, these costs are likely to increase in the years to come [19, 20]. Unfortunately, back pain in older people is poorly understood, as the vast majority of back pain research has focused on the working-aged population [21-25]. Historically, people aged > 65 years have been underrepresented in back pain research [21-23, 25] leading to a knowledge gap that might have important implications at both an individual and societal level [21]. Generalization of results from younger to older people with back pain cannot be done automatically [21]. First, older people present with more comorbidities [26] and mobility limitations [27, 28] than younger adults. Second, specific back conditions such as spinal stenosis and vertebral fractures occur more frequently in the older age groups [29-33]. These aspects might impact both the individual and societal burden of back pain.

To reduce the burden of back pain and improve healthcare-related decision-making among older people with back pain, studies focusing on older people are needed. Therefore, an international consortium, the Back Complaints in the Elders (BACE), was established in 2008 by a group of international back pain researchers in order to create a standardized methodology for conducting large cohort studies among older people with back pain [34]. Norway joined the consortium in 2015 and established the Back Complaints in the Elderly - Norway study (BACE-N) [35]. The main results in this thesis are based upon data from BACE-N.

The overall aim of this thesis was to develop new knowledge regarding cost of illness due to back pain among older people. More specifically, to describe healthcare utilization and estimate associated costs among older people seeking primary care due to back pain, and to identify and replicate modifiable prognostic factors of high costs related to healthcare utilization. Furthermore, to promote future comprehensive healthcare economic evaluations, in which costs related to productivity loss are included, we wanted to evaluate the measurement properties of the iMTA Productivity Cost Questionnaire (iPCQ), a standardized instrument for measuring health-related productivity loss. Productivity loss is often a large part of the total costs related to health and healthcare interventions [36, 37].

BACKGROUND

2.1 Back pain

2.1.1 Definition, prevalence, categorization, and current understanding

Back pain is a symptom defined by the location of pain, typically between the lower rib margins and the inferior gluteal folds [1]. Back pain is common among people of all ages [32, 38]. A systematic review of 165 prevalence studies from 54 countries reported a median (interquartile range, IQR) lifetime prevalence of low back pain of 42% (15-60%) [39]. Furthermore, it was reported that low back pain is slightly more common among females compared to males, and in those aged 40 to 80 years [39]. The prevalence of benign or mild back pain appears to decrease with increasing age, after a peak in the sixth decade [40-42]. However, the prevalence of severe and disabling back pain continues to increase with increasing age [40-42].

Back pain is often categorized according to the duration of symptoms, with acute pain having a duration of < 6 weeks, sub-acute pain of 6 to 12 weeks, and persistent pain of > 12 weeks [24, 43]. For most people with back pain, approximately 90%, it is not currently possible to accurately identify the specific cause of their symptoms, and they are therefore categorized as having non-specific back pain [25, 32, 38]. Some people with back pain, approximately 2-11% within primary care, are categorized as having neurological symptoms such as radicular pain, radiculopathy, or spinal stenosis [25, 32, 44]. Only a small proportion of people, approximately 1-5% within primary care, are categorized as having a specific pathological cause of back pain such as vertebral fractures, inflammatory disorders, spondyloarthropathy, malignancy, infections, cauda equina syndrome, and intra-abdominal causes [32].

The prevalence of spinal stenosis and specific pathological causes of back pain, such as osteoporotic vertebral fractures, infections, and malignancy, increases somewhat with age [29-33]. The prevalence of lumbar spinal stenosis has been estimated to be 9% in the general population, and up to 19-47% in people > 60 years, depending on the criteria used [45, 46]. However, imaging findings are also present in a high proportion of asymptomatic individuals [47]. The prevalence of spinal stenosis in asymptomatic people > 60 years has been estimated to be 21% [48]. The prevalence of vertebral fractures has been estimated to be 12-14% in the Norwegian general population, and 19-20% in people > 70 years [49]. A study of 2.383 patients with low back pain referred for imaging by a specialist found that the incidence of

vertebral fractures, spinal infections, and malignancy increased with age. For vertebral fractures, the incidence increased from 4 to 75 to 87 per 100.000 person-years (p-y) in the age groups of 25-34 years, 65-74 years, and 74-84 years, respectively. For spinal infection, the incidence increased from 9 to 17 to 52 per 100.000 p-y, respectively. For malignancy, the incidence increased from 4 to 34 to 78 per 100.000 p-y, respectively [31].

In recent decades, the biopsychosocial model [50] has dominated as a theoretical framework for understanding the complexity of back pain and back pain related disability [32, 38]. Multiple factors including biophysical and psychosocial factors, as well as pain-processing mechanisms and coexisting comorbidities, are important to consider in clinical work and research involving patients with back pain and related disability [50].

2.1.2 Burden of disease

The term burden of disease generally describes the total, cumulative consequences of a defined disease with respect to disability and premature mortality [3]. Two Global Burden of Disease (GBD) Study metrics commonly used to provide information on the social impact of back pain are: years lived with disability (YLDs) and disability adjusted life years (DALYs) [51]. YLD estimates the amount of healthy life that is lost due to poor health, where 1 YLD represents the equivalent of 1 full year of healthy life lost. DALYs combine years of life lost due to poor health and years of life lost due to premature mortality. One DALY represents 1 year of healthy life lost because of poor health or premature mortality [3, 51]. For back pain, the values of YLDs and DALYs will be the same, as no evidence for mortality from back pain has been found in the GBD Study [17].

The GBD Study has defined low back pain as the number one cause of years lived with disability for nearly three decades [16, 17, 52]. Globally, in 2019, 63.7 million YLDs were estimated to be caused by low back pain [17, 52]. Low back pain YLD rates were higher among females compared to males and increased with age, peaking in the age group of 80-84 years old for both sexes [17, 52]. The World Health Organisation has defined back pain as one of the most disabling conditions among older people [53].

2.1.3 Cost of illness

Back pain imposes a major socio-economic burden related to costs of healthcare, reduced work productivity, early retirement, and strain on the welfare system [18-20, 32]. Back pain is one of the most prevalent complaints encountered in primary care [54-58]. The prevalence of seeking healthcare for back pain increases with age [59-61]. In Europe, back pain is the most

common cause of medically certified sick leave and early retirement [62]. The economic impact related to back pain is comparable to other prevalent, high-cost conditions such as cardiovascular diseases [63].

Cost of illness (COI) studies aim to identify and estimate the costs of health problems, including direct, indirect, and intangible costs [8, 9]. Direct costs are monetary costs related to healthcare and non-healthcare costs such as transportation. Indirect costs are monetary costs related to productivity loss (absenteeism, presenteeism, and productivity costs related to unpaid work). Intangible costs are related to discomfort, pain, anxiety, or inconvenience, usually measured by a reduction in quality of life [7-9].

A recent systematic scoping review of 45 COI studies of back pain reported that national annual total cost estimates in 2015 USD ranged from $259 million ($29 per capita) in Sweden to $72 billion ($868 per capita) in Germany [20, 64, 65]. Direct comparisons of costs between COI studies are not feasible due to significant differences in the methodologies used. However, in studies providing estimates of both direct and indirect costs (n = 15), indirect costs far outweighed direct costs [20].

COI studies provide an important guide and resource for policy development, priority setting, and management of public health [9, 66]. To improve the use of scarce healthcare resources, and thus reduce the burden on our healthcare systems, researchers have highlighted the importance of monitoring healthcare utilization and direct costs related to back pain [43, 67]. Table 1 provides an overview of previous COI studies (n = 26) investigating direct costs related to back pain. The one-year mean total direct cost per patient estimated in 2020 euros ranged from €761 in Germany [64] to €13.783 in Japan [68], and outpatient costs (such as general practitioner (GP), physiotherapy, and chiropractor visits) seems to be the main cost driver. To the best of my knowledge, no COI study investigating direct costs related to back pain has been conducted within the Norwegian healthcare system or among a sample of exclusively older people [20]. Generalization of results between different healthcare systems cannot be done automatically, nor can it be done from younger adults to older people [20, 21]. The prevalence of seeking healthcare for back pain increases with age [59-61], and older people present with comorbidities [26], mobility limitations [27, 28], and specific pathological causes of back pain more frequently than younger adults [29-33]. Therefore, it is reasonable to expect that direct costs related to healthcare utilization are higher among older people as compared to younger adults.

The number of older people is expected to steadily increase in the years to come, thus, back pain related healthcare utilization among older people is also expected to rise substantially [60]. Consequently, monitoring healthcare utilization and direct costs among older people is an important step towards improving our use of scarce healthcare resources and addressing the global burden of back pain.

Identified research gap #1: *To the best of my knowledge, no previous study has estimated healthcare utilization and related costs among a sample of exclusively older people with back pain*

Table 1. Overview of cost of illness studies investigating the direct costs of back pain (adapted from Zemedikun et al. 2021 [20])

Lead author Design and country	Population	Main data source, year of data collection Cost components	Cost estimation	Results
Studies adopting a health system perspective				
Depont [69] Retrospective France	796 GP patients with CLBP Mean (SD) age 53 (11)	Surveys/questionnaires, 2001 *Direct costs: GP visits, investigations, medication, hospitalization, other medical and non-medical resources (e.g., home or vehicle adaptation, domestic help)*	Bottom-up	6-month mean (95% CI) total direct cost per patient was estimated at €716 (644–798) in 2007 prices. Largest cost components were physiotherapy and allied specialists (23%), medication (20%), and hospitalization (17%). > 50% of patients had ≥ 1 investigation prescribed by a GP (64% X-rays, 23% CT, 10% MRI, 3% osteodensitometry)
Hong [70] Matched case-control UK	52.986 GP patients with CLBP Age >60 (60% of sample)	UK General Practice Research Database, 2007-09 *Direct costs: GP visits, secondary care referrals, medication*	Bottom-up	12-month total direct cost for patients with CLBP were twice that of matched controls (£1074 vs £516) in 2009 prices. Of cost difference, 59% was related to GP visits, 22% to secondary care referrals, 19% to pain medication. NSAIDs and opioids were the most prescribed medications
Studies adopting a societal perspective				
Walker [71] Retrospective Australia	People diagnosed with LBP	Australian adult LBP prevalence survey and other registry, 2001 *Direct costs: diagnosis, treatment, rehabilitation (hospitalization, medical, ancillary, complementary healthcare) including prevention, research, training, administration and support-care facilities, transport, out-of-pocket costs, special clothing requirements*	Top-down	Total annual direct cost of LBP in Australian adults was estimated at AU$1.02 billion in 2001. Most costs (70%) were related to chiropractors, GPs, massage therapists, physiotherapists, and acupuncturists
Van Zundert [72] Retrospective Belgium	People with LBP	Official registrations, 1999 *Direct costs: prescription medication, rehabilitation (physiotherapy, kinesitherapy), TENS, non-surgical interventions (e.g., epidural injections), radiofrequency treatment, surgery*	Top-down	Total annual direct cost of LBP in Belgium was estimated at €187 million in 1999. Most costs were related to conservative rehabilitation (61%) and medication (19%)
Coyte [73] Retrospective Canada	People diagnosed with musculoskeletal disorder	Ontario Health Survey data, 1990-94 *Direct costs: hospitalization and other institutions, GP and other health professional visits, medication, research, other items*	Top-down	Total annual direct cost of musculoskeletal disorders in Canada was estimated at 7.5 billion CAD in 1994. Most costs were related to hospitalization (42%) and GPs (27%). Total annual direct cost of back and spine disorders was estimated at 0.7 billion CAD
Hemmila [74] Retrospective and prospective Finland	114 primary care patients with BP for ≥ 7 weeks Mean (SD) age 42 (10)	Social Insurance Institution files and patient records, 1994 *Direct costs: Official medical care and complementary therapy incl. the randomized therapies, visits to health centres and hospitals, institutional rehabilitation, medication*	Bottom-up	12-month mean total direct cost of BP was estimated at $500 in 1994 prices
Becker [75] Prospective Germany	1.211 GP patients with LBP Mean (SD) age 49 (14)	Questionnaires and interviews, 2004 *Direct costs: GP and specialist visits, diagnostic and therapeutic procedures incl. physiotherapy, medication, hospitalization and rehabilitation care, auxiliaries*	Bottom-up	6-month mean (95% CI) total direct cost for acute and chronic LBP was estimated at €456 (€366-588) and €854 (€714-1045) in 2004 prices, respectively. Largest cost component was therapeutic procedures (27-30%)
Wenig [64] Retrospective Germany	5.650 people with BP past 3 months Mean (SD) age 44 (15)	Postal survey, German Back Pain Research Network, 2003-06 *Direct costs: medication, GP visits, physiotherapy, orthopaedic aids, hospitalization, rehabilitation treatment*	Bottom-up	12-month mean (95% CI) total direct cost per patient with BP was estimated at €613 (537-706) in 2005 prices. 29% of costs were related to hospitalization, 23% to GP visits, 22% to physiotherapy, 15% to rehabilitation, 7% to medication, 5% to orthopaedic aids
Itoh [76] Retrospective Japan	Adults with work related LBP Age range 20-64	Survey Medical Care Activities in Public Health Insurance, 2011 *Direct costs: medication, laboratory tests, equipment, labour costs, etc.*	Bottom-up	Total annual direct cost of work-related LBP in Japan was estimated at 82 billion yen in 2011. Largest cost component was outpatient costs (56 billion yen)
Montgomery [68] Retrospective Japan	392 adults reporting CLBP (LBP the prior month and diagnosed by GP with LBP the prior 3 months) Mean (SD) age 54 (14)	Japan National Health & Wellbeing Survey, 2011 *Direct costs: GP visits, emergency room visits, hospitalization*	Bottom-up	12-month mean total direct cost per patient with CLBP was estimated at €12.551 in 2011 prices. Largest cost component was hospitalization (€10.927)

Table 1. (Continued)

Boonen [77] Prospective The Netherlands	110 patients with CLBP (> 6 months) referred to specialist Mean (SD) age 41 (9)	Cost diaries from three cohorts, 2002 *Direct costs: GP, specialist, physiotherapy, and psychologist visits, complementary medicine, hospitalization, rehabilitation stays, medication (prescription, over-the-counter)*	Bottom-up	12-month mean total direct cost per patient was estimated at €1.104 in 2002 prices. Largest cost component was health care provider visits (29%)
Lambeek [78] Retrospective The Netherlands	People diagnosed with BP	National registries and authorities, 2002-2007 *Direct costs: hospitalization (in- and outpatient care, medical procedures, diagnostic tests), prescription medication, GP and allied health care visits (physical and exercise therapy, manual therapy)*	Top-down	Total annual direct cost of BP in the Netherlands was estimated at €421-479 million in 2002-07. Largest cost component was allied health care visits (€232-283 million)
van Tulder [79] Retrospective The Netherlands	People diagnosed with BP	Survey and registry data, 1991 *Direct costs: hospitalization (clinical care treatment, examinations, medication, paramedical care, operating room), medical specialist, GP visits (estimated to 10% of costs of hospitalization), paramedical care (physiotherapy). Out-of-pocket/civil services cost not included*	Top-down	Total annual direct cost of BP in the Netherlands was estimated at $368 million in 1991. 57% of cost was related to hospitalization, 36% to paramedical care, 6% to GPs, 1% to medical specialists
Alonso-Garcia [80] Retrospective Spain	4.826 people reported to suffer from LBP in the last 12 months Age ≤65 (65% of sample)	National Health Survey, 2017 *Direct costs: GP and specialist visits, emergency room, diagnostic tests, hospitalization, physiotherapy, psychologists, and medication (analgesics, anti-rheumatics)*	Bottom-up	Total annual direct cost of LBP in Spain was estimated at €2280 million in 2017. Largest cost component was specialist visits (€948 million). Total annual direct cost per case was estimated at €279
Ekman [65] Retrospective Sweden	People diagnosed with BP	Survey and registry data, 2001 *Direct costs: hospitalization, outpatient visits, physiotherapy, prescription medication*	Top-down	Total annual direct cost of LBP in Sweden was estimated at €308 million in 2001. Largest cost component was physiotherapy (€170 million)
Ekman [81] Retrospective Sweden	302 GP patients with LBP ≥ 50% of days past 3 months Mean (SD) age 49 (14)	Survey/questionnaire, 2002 *Direct costs: hospitalization, outpatient care, diagnostic tests, medication, orthopaedic aids, physiotherapy, chiropractors, paid home help*	Bottom-up	12-month mean total direct cost per patient with chronic LBP was estimated at €3.100 in 2002 prices
Hansson [82] Prospective Sweden	1.146 employees with full work incapacity ≥ 28 days due to GP-certified LBP or neck problems Age range 18-59	Prospectively entered diaries and questionnaires, 1994-95 *Direct costs: examination, treatment, rehabilitation. Transportation not included*	Bottom-up	24-month mean total direct cost per subject was estimated at €1.810 in 1995 prices
Jonsson [83] Retrospective Sweden	People diagnosed with BP	National Board of Health and Welfare's register, 1994 *Direct costs: primary care, in- and outpatient care, institutional care, prescription medication*	Top-down	Total annual direct cost of BP in Sweden was estimated at 832 million SEK in 1994. Largest cost components were hospital visits (224 million SEK), short-term inpatient care (195 million SEK), and primary health care (181 million SEK)
Olafsson [84] Retrospective Sweden	129.973 patients with registered LBP (or radiating leg pain) healthcare visit or sick leave/early retirement Mean age 53	Administrative database VEGA and PAR, 2008-11 *Direct costs: healthcare visits (GPs, other physicians, nurses, physiotherapists, chiropractors, psychologists, other healthcare staff), prescription medication (pain and depression medication, muscle relaxants, anti-inflammatory)*	Bottom-up	Mean (SD) total direct cost per episode was estimated at €917 in 2016 prices. Largest cost component was GP visits (34%)
Wieser [85] Retrospective and prospective Switzerland	1.253 adults with LBP ≥ 4 weeks Mean age 52	Large population-based survey and cost diary, 2005 *Direct costs: GP and specialist visits, different types of therapy, pain medication, surgery, hospitalization, utility devices, physical training (treatment, prevention), ADL assistance*	Bottom-up	12-month mean total direct cost for an individual reporting LBP ≥ 4 weeks was estimated at €1.842 in 2005 prices. Total annual direct cost of LBP in Switzerland was estimated at €2.6 billion in 2005. Largest cost components were ADL assistance (18%) and physiotherapy (11%)
Icatasiotlu [86] Retrospective Turkey	662 patients consulting physical medicine or rehabilitation specialists with CLBP (LBP ≥ 50% of days past 3 months) Mean (SD) age 46 (15)	Surveys/questionnaires, 2013 *Direct costs: hospitalization, outpatient clinic visits, diagnostic tests, medication, orthopaedics aids, physiotherapy, home care financial support*	Bottom-up	Total annual direct cost of CLBP in Turkey was estimated at 714.735 TL in 2013. Total annual direct cost per case was estimated at 1080 TL

Table 1. (Continued)

Yumusakhuylu [87] Retrospective Turkey	211 patients diagnosed with CLBP (pain ≥ 50% of days past 3 months) Mean age 45	Surveys/questionnaires, 2011 *Direct costs: medical visits, diagnostic tests, medication, hospitalization, orthopaedic aids, physiotherapy, housing benefit*	Bottom-up	12-month mean direct costs per patient with CLBP were estimated at €346 in 2011 prices. Most prescribed medication was NSAIDs. 30% of patients received radiography, 52% MRI, 3% CT, 9% bone mineral density tests, 10% laboratory tests (CBC, ESR, and CRP)
Maniadakis [63] Retrospective UK	People diagnosed with BP (between inferior angle of scapula and gluteal folds)	Office of Population Censuses and Surveys, 1998 *Direct costs: GP, private consultant, physiotherapy, osteopath, chiropractor, and other specialist visits, hospitalization, outpatient visits, emergency room, inpatient days, imaging, medication (prescription and over-the-counter), community health and social services*	Top-down	Total annual direct cost of BP in the UK was estimated at £1.632 million in 1998. 37% of costs were related to physiotherapy and allied specialists, 31% to the hospital sector, 14% to primary care, 7% to medication, 6% to community care, 5% to imaging
Luo [88] Matched case-control USA	25.9 million adults with BP (upper/lower part) related diagnostic code Mean age 48	Medical Expenditure Panel Survey, 1998 *Direct costs: in- and outpatient care, office-based visits, emergency room, prescription medication, home health services, dental care, vision aids, medical equipment purchase. Incl. GP, GP assistant, chiropractor, physiotherapist, occupational therapist, psychologist, nurse, social worker visits. Nursing home care not included*	Bottom-up	Total annual direct cost of BP in the USA was estimated at $90.7 billion in 1998. Largest cost components were inpatient care (31%) and office-based visits (26%). The top quantile accounts for ≥ 75% of costs. On average, adults with BP had 60% higher costs than adults without BP ($3.498 vs $2.178)
Martin [89] Matched case-control USA	1997 sample: 3.139 adults with spine problems (cervical-lumbar) and 19.906 adults with no spine problems. 2005 sample: 3.187 adults with spine problems (cervical-lumbar) and 19.071 adults with no spine problems Mean age 44-49	Medical Expenditure Panel Survey, 1997-2005 *Direct costs: in- and outpatient, emergency room, prescription medication*	Bottom-up	In 1997, mean age- and sex adjusted cost for adults with spine problems was $4.695, compared with $2.731 among those without spine problems (2005 prices). In 2005, mean age- and sex-adjusted cost for adults with spine problems was $6.096, compared with $3.516 among those without spine problems. 36% of cost difference in 2005 was accounted for by outpatient services
Smith [90] Retrospective USA	12.104 adults reported being "bothered" by BP or had a BP disability day. 3.842 were categorized as having chronic BP (bothered by BP or had a BP disability day in ≥ 3 survey rounds) Mean age 48-52	Medical Expenditure Panel Survey, 2000-07 *Direct costs: office-based visits, outpatient visits, emergency room, prescription medication*	Bottom-up	Total 24-month direct cost of chronic BP and nonchronic BP in the USA was estimated at $36 and $17 billion in 2006-07 (2010 prices), respectively. Mean 24-month direct cost per patient with chronic BP and nonchronic BP was estimated at $3.152 and $903, respectively. The top quantile accounts for 72% of costs

ADL indicates activities of daily living; AU$, Australian dollar; BP, back pain; CAD, Canadian dollar; CI, confidence interval; CLBP, chronic low back pain; GP, general practitioner; LBP, low back pain; SEK, Swedish krona; TL, Turkish lira; €, euro; £, British pound; $, United States dollar. The original paper of Zemedikun [20] presented the main characteristic of 45 studies. In this table, only studies including direct costs and adopting a health system perspective (include only direct costs) or a social perspective (including direct and indirect costs) are presented. Studies adopting an insurer or employer perspective were omitted.

2.1.4 Clinical guidelines and management

Updated international clinical guidelines provide more or less consistent recommendations for how to assess and treat adults with back pain. First-line treatment should include advice to remain active, education, and reassurance, as well as exercise and cognitive behavioural therapy for those with persistent symptoms. Adjunctive options in case of an inadequate response to first-line treatment could be manual therapy, acupuncture, pharmacological treatment, and optional surgery [43, 91-93]. Laboratory tests and imaging should not be routinely used, but rather be reserved for patients for whom the result is likely to alter management [43, 91-94]. Finally, a key recommendation is to adopt a stratified healthcare approach guided by the patient's response to care or the results of risk prediction tools (such as the StarT Back Screening Tool [95]) [43, 96, 97]. Targeting the use of resources to those most likely to benefit from them might allow for an improvement in patient outcomes while reducing avoidable costs and the burden on our healthcare systems [43, 97, 98].

The evidence underlying these guidelines is mainly based on research conducted on middle-aged people with low back pain, and whether they are appropriate for use among older people and people with thoracic back pain is not known [21, 43, 92, 99]. There are no specially made guidelines for older people with back pain and guidelines for thoracic back pain are lacking [92]. Therefore, clinicians have to use the above-mentioned guidelines, among older people with back pain and people with thoracic back pain.

Although these guidelines are well established and health providers report being aware of them, concerns about substantial gaps between guidelines and practice have been highlighted [67, 96]. Problems include underuse of high-value care (e.g., education, advice to remain active and exercise), overuse of low-value care (e.g., pharmacological treatment as first-line treatment and high imaging rates), and thereby a misuse of limited healthcare resources [32, 43, 67, 91]. The extent to which this concern also applies to older people seeking primary care due to back pain is not known. As mentioned above, monitoring healthcare utilization and related costs among older people is an important step towards improving our use of scarce healthcare resources and addressing the global burden of back pain [43, 67]. Undoubtedly, it is also essential to discuss healthcare utilization and related costs in the context of clinical guidelines [43, 67].

2.1.5 Modifiable prognostic factors of high costs related to healthcare utilization

It is well known that most healthcare utilization and its related costs stem from a relatively small group of patients with back pain [100], and, more importantly, that many of these patients receive unnecessary, ineffective, and in some cases, harmful treatment [67, 96]. In order to reduce the use of low-value care and its associated costs, knowledge regarding modifiable prognostic factors associated with high costs related to healthcare utilization is needed. Information about modifiable prognostic factors can inform development of targeted interventions, which may enhance the clinical effectiveness and cost-benefits. Cost-effective interventions are crucial in order to improve the use of scarce healthcare resources and reduce the economic burden on our healthcare systems [12].

As described by Andersen's Behavioral Model of Health Services Use [101], healthcare utilization is a function of people's predisposition to use services, factors that enable or impede use, and the need for care. To the best of my knowledge, only four prospective studies have explored modifiable prognostic factors associated with costs or high costs related to healthcare utilization among patients with back pain [61, 75, 102, 103] (Table 2), of which one explored costs related to healthcare utilization as part of societal costs [102]. The four studies included mainly middle-aged patients with back pain from primary and secondary care settings in Germany [75], the Netherlands [102], and the USA [61, 103]. Potential modifiable prognostic factors identified in these studies were: pain severity [75, 102, 103], impact of pain experience [102], radiating pain [75], disability [75, 102, 103], comorbidity [61], quality of life [75, 102], physical health [102], fear avoidance [75], and depression symptoms [75]. In addition to the four prospective studies, some retrospective studies have been published [64, 104-106], as well as studies on related topics (healthcare seeking, numbers of healthcare contacts, healthcare utilization patterns) [107-109], and studies on patients with musculoskeletal disorders [110-112]. Potential modifiable prognostic factors identified in these studies were: pain severity [64, 104-106, 109-112], disability [64, 104, 105, 109, 111, 112], comorbidity [105], health-related quality of life [110], back beliefs [104], fear avoidance [104, 107], stress [106, 107, 111], anxiety [108], and depression symptoms [105, 108, 112].

Patients with high costs related to healthcare utilization are a diverse population, which seems to vary across different health problems, provider characteristics, payer types, countries, and age groups [108, 113]. Therefore, generalization of results cannot be done automatically between different health problems and settings, or from younger to older people

with back pain [21]. Identifying modifiable prognostic factors of high costs related to healthcare utilization among older people is an important step towards addressing the global burden of back pain and decreasing waste of valuable healthcare resources [12, 38, 67, 96].

Identified research gap #2: *To the best of my knowledge, no previous study has identified modifiable prognostic factors of high costs related to healthcare utilization among a sample of exclusively older people with back pain*

Table 2. Overview of prospective studies investigating associations between modifiable prognostic factors and (high) costs related to healthcare utilization among patients with back pain

Lead author Country	Start point	Main data source, year of data collection Cost components	Endpoint	Results
Becker [75] Germany	1.211 primary care GP patients with LBP Mean (SD) age 49 (14)	Questionnaires and interviews, 2004 *Direct costs: GP and specialist visits, diagnostic and therapeutic procedures incl. physiotherapy, medication, hospitalization and rehabilitation care, auxiliaries*	High costs related to healthcare utilization (cost ≥ €983)	High direct costs were associated with depression symptoms (assessed by the CES-D, score > 23) (adjusted OR 1.81, 95%CI 1.11-2.98) and pain severity/disability (von Korff's severity of chronic pain scale) (adjusted OR range 1.42-4.78). Fear avoidance beliefs (FABQ), quality of life (EuroQol), and radiating pain to the leg were of minor importance for high direct costs
Engel [103] USA	1.059 primary care patients with BP Age 18-64 (84% of sample)	Group Health Cooperative of Puget Sound data, 1989-90 *Direct costs: primary care and specialist visits, hospitalization, radiographic imaging, pain medication*	High costs related to healthcare utilization, top 21st percentile (cost ≥ $600)	The most consistent and strongest prognostic factor of high direct costs was pain intensity and related dysfunction (assessed by chronic pain grade) (adjusted OR range 1.4-7.2). Depression symptoms (SCL-90) were not associated with high direct costs
Ritzwoller [61] USA	16.567 primary and specialist care patients with LBP Age 18-64 (79% of sample)	Keiser Permanente Colorado claims database, 1996-2001 *Direct costs: outpatient care (primary and specialty care, physiotherapy, mental and behavioural health), hospital-based care (inpatient stays, emergency room visits, visit and observation stays), medication, radiographic imaging*	Costs related to healthcare utilization	Total annualized direct costs increased with physical and mental health comorbidities
Mutubuki [102] The Netherlands	6.316 patients with CLBP (>3 months) who showed no improvement after conservative treatment, referred to a pain clinic Mean (SD) age 57 (13)	Questionnaires, 2013-15 *Direct costs: GP visits, manual therapy, physiotherapy, exercise therapy, secondary care (diagnostic/therapeutic interventions), hospitalization*	High societal costs (costs related to healthcare utilization and absenteeism), top 10th percentile (cost ≥ €11.922)	High societal costs were associated with poor physical health (assessed by Rand-36, 0-100) (OR 0.93, 95%CI 0.89-0.97), high functional disability (ODI, 0-100) (OR 1.04, 95%CI 1.02-1.05), decreasing pain (NPRS, 0-100) (OR 0.99, 95%CI 0.98-0.99), high impact of pain experience (MPI interference, 0-100) (OR 1.02, 95%CI 1.01-1.03), and low health-related quality of life (EQ-5D-3L, 0-100) (OR 0.99, 95%CI 0.99-1.00). Patient expectation of improvement after treatment was not associated with high societal costs

BP indicates back pain; CES-D, The Center for Epidemiologic Studies Depression Scale; CI, confidence interval; CLBP, chronic low back pain; EQ-5D-3L, EQ-5D-3L; FABQ, The Fear Avoidance Beliefs Questionnaire; GP, general practitioner; LBP, low back pain; MPI, Multidimensional Pain Inventory; NPRS, Numeric Pain Rating Scale; ODI, Oswestry Disability Index; OR, odds ratio; Rand-36, Rand-36; RCT, randomized controlled trial; SCL-90, The Symptoms Checklist-90 Revised.

2.2 Measurement of productivity costs

Productivity costs often reflect a large part of total costs related to health and healthcare interventions [36, 37], and the majority of guidelines recommend that these costs be included in healthcare economic evaluations [114]. Nevertheless, productivity costs are often omitted from these evaluations [115], an oversight that might be explained in part by the lack of standardized methodologies used to measure and value productivity costs [115, 116]. Currently, there is no gold standard method for measuring productivity costs [20, 37, 115, 117]. Nevertheless, there is a general agreement that one should not only measure productivity costs related to absenteeism and presenteeism, but also costs related to unpaid work such as household work, care work, and volunteer work [10]. Several instruments measuring productivity costs are available, but the majority are not specifically intended or suited for use in economic evaluations [10, 37, 118-120]. Researchers have therefore highlighted a need for valid instruments suited for use in economic evaluations [20, 37, 118].

The iMTA Productivity Cost Questionnaire (iPCQ) has recently been developed to cover the three domains of productivity costs [10]. It was designed to optimize features of existing instruments, and to be a short, generic, patient-reported outcome measure which allows for the quantification and valuation of all productivity costs (absenteeism, presenteeism, and costs related to unpaid work) [10]. The iPCQ is assumed to be a promising instrument [37], and a recent systematic review recommended it as probably the most suitable instrument for use in health economic evaluations [121]. However, there is no Norwegian version of the iPCQ, and only two studies have tested its measurement properties. Bouwmans et al. [10] confirmed its feasibility and face validity. Furthermore, in a modified version (iPCQ-VR), Beemster et al. [122] tested the reliability, agreement, and responsiveness of the core parts of absenteeism and presenteeism; they found good measurement properties on long-term absenteeism and poor measurement properties on short-term absenteeism and presenteeism. The iPCQ should be translated and cross-culturally adapted into Norwegian. Moreover, further validation of the iPCQ is required [10] and expected to improve the toolset needed to conduct a comprehensive healthcare economic evaluation.

Identified research gap #3: *To the best of my knowledge, the iPCQ has not been tested with respect to content, construct, or criterion validity, or reliability. Furthermore, there is no Norwegian version of this instrument*

AIMS

The primary aim of this thesis was to develop new knowledge on the cost of illness due to back pain among older people. More specifically, to describe healthcare utilization and associated costs among older people seeking primary care due to back pain and identify modifiable prognostic factors of high costs related to healthcare utilization. A secondary aim was to evaluate the measurement properties of the iMTA Productivity Cost Questionnaire (iPCQ). The specific aims were:

I. To describe healthcare utilization and estimate associated costs during one-year of follow-up among older people seeking primary care due to a new episode of back pain, and to describe healthcare utilization across patients with different risk profiles (**Paper I**)

II. To identify modifiable prognostic factors for high costs related to healthcare utilization among older people seeking primary care with a new episode of back pain, and to replicate the identified associations of modifiable prognostic factors in a similar cohort of older patients with back pain (**Paper II**)

III. To translate and cross-culturally adapt the original iPCQ into Norwegian and to test its measurement properties among patients with musculoskeletal disorders (**Paper III**)

IV. To evaluate the criterion validity of the iPCQ by comparing iPCQ-reported occurrence and duration of long-term absenteeism (> 4 weeks) with public registry data among people on sick leave due to musculoskeletal disorders (**Paper IV)**

METHODS

This thesis is built upon four scientific papers: two prospective cohort studies (Papers I and II) and two methodological studies (Papers III and IV). It is based on data from the Back Complaints in the Elderly - Norway study (BACE-N) [35], a prospective observational cohort study within Norwegian primary care. In addition, data from two other research projects have been used: 1) the Back Complaints in the Elders study (BACE-D) [34], a prospective observational cohort study within Dutch primary care, and 2) the MI-NAV project [123], a large-scale project including people on sick leave due to musculoskeletal disorders conducted within the Norwegian Labour and Welfare Administration. Furthermore, data from a sub-study of BACE-N has been used in which measurement tools were validated, a cross-sectional study including a test-retest assessment of patients with musculoskeletal disorders within Norwegian secondary care. A method overview of the four papers is given in Table 3.

The PROGnosis RESearch Strategy (PROGRESS) framework [12, 13] was used as a theoretical framework for Papers I and II. Paper I is considered to be part of overall prognosis research. Paper II is considered to be part of prognostic factor research i.e., identification of prognostic factors, including external replication. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) was used as a theoretical framework for Papers III and IV [124, 125].

Table 3. Overview of study design, recruitment, study population, main outcomes, and data analysis

	Paper I	Paper II	Paper III	Paper IV
Study design	Cohort study One-year follow-up	Cohort study One-year follow-up	Translation and cross-cultural adaptation Cross-sectional study including a test-retest assessment	Cross-sectional study including public registry data, one-year retrospective
Recruitment	Recruitment from physiotherapists, chiropractors, and GPs working in Norwegian primary care as part of BACE-N [35]	Recruitment from physiotherapists, chiropractors, and GPs working in Norwegian primary care, as part of BACE-N [35], and GPs working in Dutch primary care, as part of BACE-D [34]	Recruitment from Norwegian secondary care at an outpatient rehabilitation clinic	Recruitment from individual profile page on the Norwegian Labour and Welfare Administration website, as part the MI-NAV project [123]
Study population	452 patients with back pain, aged ≥ 55	452 patients with back pain, aged ≥ 55, and 675 patients with back pain, aged >55	115 patients with musculoskeletal disorders, aged ≥ 18, working or on sick leave	144 participants with musculoskeletal disorders, aged ≥ 18, on sick leave for ≥ 4 weeks
Main outcomes	Healthcare utilization and related costs	Costs related to healthcare utilization	iMTA Productivity Cost Questionnaire	iMTA Productivity Cost Questionnaire
Data analysis	Descriptive and between-group analysis	Binary logistic regression analysis	Content and construct validity, and test-retest reliability analysis	Criterion validity analysis

BACE-D indicates Back Complaints in the Elders study; BACE-N, Back Complaints in the Elderly - Norway study; iMTA, Institute for Medical Technology Assessment; GP, general practitioner.

3.1 Ethical approval and considerations

All papers within this thesis adhered to the Code of Ethics of the World Medical Association (Declaration of Helsinki). The Norwegian Regional Committee for Medical Research Ethics classified Papers I-IV as quality assessment studies (BACE-N, ref no. 2014/1634/REK vest; MI-NAV, ref no. 2018/1326/REK sør-øst A), and specified that a quality assessment study does not require their explicit approval. The Medical Ethics Committee of the Erasmus Medical Center, the Netherlands approved Paper II (BACE-D, ref no. NL24829.078.08). The Norwegian Social Science Data Service (NSD) approved Papers I-IV (BACE-N, ref no. 42149; MI-NAV, ref no. 861249). Patient representatives were part of the scientific board of the BACE-N and the MI-NAV project and were involved in designing and establishing the studies.

All participants included in Papers I-IV signed a written consent form prior to inclusion and were informed that they could withdraw at any time. All data used and analysed in Papers I-IV were stored in a secure IT platform. Data included in Papers I, II, and IV were stored at the Services for sensitivity data (TSD). Data included in Paper III were stored in a safe environment within the Oslo Metropolitan University network. In addition, data included in Paper II (BACE-D material) were stored in a safe environment within the Erasmus MC network.

Papers I, II, and III were funded by Oslo Metropolitan University, the Norwegian Fund for Post-Graduate Training in Physiotherapy (grant no. 90749), and “Et liv i bevegelse” (A life in movement) - the Norwegian chiropractors’ research foundation. Paper IV was funded by Oslo Metropolitan University and the Research Council of Norway. The funding organizations had no part in the planning, performing, or reporting of the papers. No authors declared any conflicts of interest with regards to Papers I-IV.

3.2 Protocol and registration

The four papers within this thesis are based on data from BACE-N [35], BACE-D [34], and the MI-NAV project [123]. The BACE-N and the MI-NAV project have been registered at ClinicalTrials.gov (NCT04261309 and NCT04196634, respectively). A study protocol of BACE-D has been published [34]. Furthermore, a statistical analysis plan was published for Papers I [126] and II [127].

3.3 Study design

Papers I and II were conducted using a prospective observational cohort study design with one-year of follow-up. Paper III was carried out in two steps. First, the original version of the iPCQ was translated and cross-culturally adapted into Norwegian. Next, the measurement properties of the Norwegian iPCQ were tested using a cross-sectional design including a test-retest assessment after 2 to 3 days. Paper IV was conducted using a cross-sectional study design on baseline data from a prospective observational cohort study and public registry data in the period from baseline to 12 months retrospectively [123].

3.4 Translation and cross-cultural adaptation (Paper III)

Translation and cross-cultural adaptation of the iPCQ were carried out according to international guidelines [128, 129]. Two native Norwegian speakers (1 philologist and 1 clinician) independently translated the original iPCQ from English into Norwegian. The two Norwegian versions were then synthesized into one before being back-translated into English. Two native English speakers (1 philologist and 1 clinician), both blinded to the original iPCQ, independently performed the back-translation and synthesized the two English versions into one. An expert committee consisting of translators and two researchers from our research group reviewed all translations. In a formal meeting, the committee discussed deviations until consensus on a prefinal version was reached. The goal was that the prefinal Norwegian iPCQ should be as concise and easy to understand as possible. The prefinal version was tested among 10 patients with musculoskeletal disorders. None of the patients had difficulty understanding the meaning of items or responses, and they found it easy to comprehend. No changes had to be made, thus, the final version of the Norwegian iPCQ evaluated in Paper III is the same as the prefinal version.

3.5 Eligibility criteria and recruitment procedures

Eligible participants for BACE-N (Papers I and II) were people 55 years of age or older seeking Norwegian primary care (physiotherapists, chiropractors, or GPs) with a new episode of back pain. Back pain was defined as pain located in the region from the top of the scapula to the first sacral vertebra. A new episode was defined as not having received primary healthcare for the same complaint in the preceding 6 months. Exclusion criteria were the inability to complete the questionnaires (because of language barriers or a cognitive disorder) or the physical examination (e.g., being wheelchair-bound). Participants within BACE-N were prospectively recruited from primary care (physiotherapists, chiropractors, or GPs) in

urban and rural parts of Norway between April 2015 and February 2020. Recruiting primary care providers were asked to inform all eligible patients about the study, to obtain the patient's permission for the researchers to contact them, and to send the patient's contact information to the researchers. We then contacted the patients by telephone, screened for eligibility, informed them about the study, answered any patient questions, and made an appointment for inclusion at one of our test stations. To facilitate recruitment, advertisements were placed in local newspapers.

Eligible participants for BACE-D (Paper II) were people over 55 years of age seeking Dutch primary care (GP) with a new episode of back pain. Back pain was defined as pain located in the region from the top of the scapula to the first sacral vertebra. A new episode was defined as not having received healthcare from a GP for the same complaint in the preceding 6 months. Exclusion criteria were the inability to complete the questionnaires (because of language barriers or a cognitive disorder) or the physical examination (e.g., being wheelchair-bound). Participants within BACE-D were prospectively recruited from Dutch primary care (GPs) in and around Rotterdam between March 2009 and September 2011. Recruiting GPs were asked to inform all eligible patients about the study, to obtain the patient's permission for the researchers to contact them, and to send the patient's contact information to the researchers. The researchers then contacted the patients by telephone, screened for eligibility, informed them about the study, answered any patient questions, and made an appointment for inclusion. Additionally, eligible patients were also retrospectively identified through the medical patient records of the recruiting GPs within 2 weeks after their GP consultation. A research assistant identified potentially eligible patients using the ICPC codes L02, L03, L84, L85, and L86. The GPs then selected eligible patients, and if a patient had visited the GP no more than 2 weeks before the search, they received a letter from the GP with an invitation to join the study and the researchers' contact information.

Eligible participants for Paper III were patients with musculoskeletal disorders aged 18 or above who were working or on sick leave. Exclusion criteria were the inability to speak, read, or write in Norwegian. Participants in Paper III were prospectively recruited from secondary care at an outpatient rehabilitation clinic (Unicare Friskvern) in Akershus, Norway between November 2015 and January 2018. The inclusion was performed by clinicians, primarily physiotherapists, who met potential participants at the clinic. Eligible participants were verbally invited to participate on an individual basis and received written and oral information about the study and the procedures involved.

Eligible participants for Paper IV were people aged 18 or above on sick leave for at least 4 weeks due to musculoskeletal disorders. Exclusion criteria were the inability to read or write in Norwegian or English, a sick leave period longer than 12 months retrospectively from baseline, and not having a reported start date of sick leave with the iPCQ. Participants in Paper IV were prospectively recruited electronically through the Norwegian Labour and Welfare Administration (NAV) website between November 2018 and Mars 2019. Eligible participants were invited to participate and received written information about the study and the procedures involved through a link on their individual profile pages on the NAV website.

3.6 Data collection

All participants included in BACE-N (Papers I and II) completed a comprehensive baseline questionnaire and went through a standardized physical examination conducted by trained local research assistants at test stations established within each recruiting area. Follow-up questionnaires were sent at 3, 6, and 12 months after inclusion. All questionnaires were preferably completed electronically through Infopad (a secure solution for data collection approved by the NSD), but paper versions were available for participants who were unfamiliar with electronic data collection. Up to two reminders were sent by text message to those who did not respond to the questionnaires.

All included participants in BACE-D (Paper II) completed a comprehensive baseline questionnaire and went through a standardized physical examination conducted by trained local research assistants at test stations established within each recruiting area. Follow-up questionnaires were sent at 3, 6, 9, and 12 months after inclusion. All questionnaires were preferably completed electronically, but paper versions were available for participants who were unfamiliar with electronic data collection.

For Paper III, all included participants completed the iPCQ as part of a comprehensive questionnaire at baseline. Furthermore, participants consenting to participate in the retest assessment completed the iPCQ at their next attendance at the clinic, preferably with a 2- to 3-day interval. All questionnaires in Paper III were completed on paper at the clinic.

For Paper IV, all included participants completed the iPCQ electronically as part of a comprehensive baseline questionnaire. Furthermore, public registry data were collected retrospectively from NAV in the period from baseline to 12 months.

3.7 Measurements

Data in Papers I-IV were collected through self-reported questionnaires and public registry. Table 4 provides an overview of the variables included in the four papers. The variables are also described in more detail below.

Table 4. Overview of variables used in Papers I-IV

	Paper I	Paper II	Paper III	Paper IV
Self-reported data				
Outcome				
Healthcare utilization	X	X		
iMTA Productivity Cost Questionnaire			X	X
Potential modifiable prognostic factors, potential covariates, other variables				
Sex	X	X	X	X
Age	X	X	X	X
Ethnicity	X	X	X	X
Education level	X	X	X	X
Employment status		X		
iMTA Productivity Cost Questionnaire	X			
First healthcare provider	X	X		
Pain location	X	X		
Radiating pain below the knee	X	X		
McGill pain drawing			X	
Numeric Rating Scale	X	X	X	X
Pain duration	X	X	X	
Pain history	X	X		
Roland-Morris Disability Questionnaire	X	X		
Self-administered Comorbidity Questionnaire	X	X		
Short-Form Health Survey 36-item	X	X	X	
Center for Epidemiological Studies - Depression questionnaire	X	X		
Fear Avoidance Belief Questionnaire - Physical Activity subscale	X	X		
The Keele STarT Back Screening Tool	X			
Expectation of recovery		X		
Physical Workload Questionnaire			X	
General Nordic Questionnaire for Psychological and Social Factors at Work			X	
Red flags	X			
Healthcare utilization prior to inclusion	X	X		
Public registry data				
Long-term absenteeism (> 4 weeks)				X
Diagnostic code related to long-term absenteeism				X

iMTA indicates Institute for Medical Technology Assessment.

Outcomes

In Paper I, the primary outcome was total costs of healthcare utilization aggregated for one year of follow-up. The secondary outcomes included the following components of healthcare utilization aggregated for one year of follow-up: number of primary care consultations, number of patients using back medication, number of patients receiving imaging (X-ray, MRI,

CT), and number of patients receiving secondary care (back operation, hospitalization, rehabilitation stay).

In Paper II, the outcome was costs related to healthcare utilization aggregated for one-year of follow-up and dichotomized as high and low. Having high costs related to healthcare utilization was defined as patients with costs in the top 25th percentile [75, 103].

Healthcare utilization in Papers I (BACE-N) and II (BACE-N and BACE-D) included: consultations with healthcare professionals (type and frequency), use of back medication (prescription and over-the-counter, type and frequency), number of diagnostic examinations (blood samples and imaging, type and frequency), number of days of hospitalization and/or rehabilitation stay (within BACE-N) and back operations. Within BACE-N, consultations with healthcare professionals and use of back medication were reported with a 3-month recall period at each timepoint of follow-up. Numbers of diagnostic examinations and days of hospitalization and/or rehabilitation stay were reported with a 3-month recall period at 3- and 6-months follow-up, and a 6-month recall period at 12-months follow-up. Back operations were reported with a 12-month recall period at 12-months follow-up. Within BACE-D, all variables, except back operations, were reported with a 3-month recall period at each timepoint of follow-up. Back operations were reported with a 12-month recall period at 12-months follow-up. Total costs of healthcare utilization per patient were estimated by multiplying frequency of use by unit costs collected from national pricelists (see section 3.8 for a detailed description of cost estimation).

In Papers III and IV, the outcome was the iMTA Productivity Cost Questionnaire (iPCQ) [130]. The iPCQ consists of 18 items and adopts a recall period of 4 weeks (except for items no. 5 and 6). In the introduction, nine items (no. A1-6 and 1-3) assess the date of reply and the following sociodemographic factors: age, sex, education level, work status, paid or unpaid work, profession, number of workdays, and work hours per week of paid work. Further, productivity costs are measured in three separate index scores with individual sum scores: “absenteeism”, with a distinction between short (≤ 4 weeks) and long-term (> 4 weeks) absenteeism, “presenteeism”, and “productivity loss in unpaid work”. To calculate productivity costs, eight core items are used. The value of absenteeism is calculated from items no. 2 (*no. of weekly work hours*), 3 (*no. of weekly workdays*), 4 (*no. of days absenteeism short term*), and 6 (*no. of days absenteeism long term*); presenteeism from items no. 2, 3, 8 (*no. of workdays with disability*), and 9 (*effective score completed work*); and unpaid work productivity loss from items no. 11 (*no. of days less unpaid work*) and 12 (*no. of hours less*

unpaid work) [130]. Productivity costs are valued in hours and can therefore be translated by standard cost prices of productivity per hour. The English and the Norwegian versions as well as the manual for the iPCQ are available from the Institute for Medical Technology Assessment (iMTA) at Erasmus University Rotterdam [131, 132].

Potential modifiable prognostic factors, potential covariates, and other variables

In order to describe the study populations in Papers I-IV, to identify modifiable prognostic factors of high costs related to healthcare utilization, and to test the measurement properties of the iPCQ, the following self-reported variables were included and measured at baseline:

- Sex (female/male)
- Age (years)
- Ethnicity
- Educational level measured as the highest education completed and categorized into low (elementary and high school level) or high (university level)
- Employment status measured by the question *"Do you have a paying job?"* and categorized into yes/no
- Productivity loss (absenteeism, presenteeism, productivity loss in unpaid work) measured by the Norwegian version of the iPCQ [130]
- First healthcare provider (physiotherapist, chiropractor, or GP)
- Pain location (thoracic, lumbar/sacral) (Papers I and II) measured by the Norwegian version [133] of the McGill pain drawing [134] (Papers III and IV)
- Radiating pain below the knee measured by the question *"Did your back pain radiate to your legs last week? If yes, how far down did the pain radiate?"* and categorized into yes/no
- Average pain severity last week or last two weeks measured by the Numeric Rating Scale (NRS) [135]. The NRS (range 0-10, higher score indicates higher pain severity) has been widely used to evaluate pain, and evidence supports its reliability and validity across many populations [136]. The NRS has proven to be preferable when examining patients with back pain [137], including for Norwegian patients [138]
- Pain duration measured by the questions: *"How many days have you had your current back pain?"* and categorized into < 6 weeks, 6 weeks to 3 months, or > 3 months (Papers I and II), and *"How many days have you had your current pain?"* (Papers III and IV)

- Pain history measured by the question *"Have you had back pain before?"* and categorized into yes/no
- Back-related disability measured by the Norwegian version of the Roland Morris Disability Questionnaire (RMDQ) [139]. The RMDQ (range 0-24, higher score indicates higher degree of back-related disability) is a widely used back-specific patient-reported measure of disability and evidence supports its validity [140], also when the Norwegian version is used [141]
- Number of comorbidities measured by the Norwegian version of the Self-Administered Comorbidity Questionnaire (SCQ) [142]. The SCQ (range 0-15, higher score indicates more comorbidities) measure 13 pre-defined comorbidities (heart disease, high blood pressure, lung disease, diabetes, ulcer or stomach disease, kidney disease, liver disease, anaemia or other blood disease, cancer, depression, osteoarthritis, back pain, and rheumatoid arthritis) and two optional comorbidities, and evidence supports its validity among people with back pain [143]. In BACE-N and BACE-D item no. 12 (back pain) was replaced with a third optional comorbidity
- Health-related quality of life measured by the physical function and mental summary score of the Norwegian version of the Short-Form Health Survey 36-item (SF36) [144]. The SF36 (range 0-100, higher score indicates better health-related quality of life) has been widely used as a generic instrument for measuring health status in various diseases and conditions [145], and has been suggested to be the most appropriate generic instrument for use in musculoskeletal conditions [146]
- Emotional well-being measured by the Norwegian version of the Center for Epidemiological Studies-Depression questionnaire (CES-D) [147]. The CES-D (range 0-60, higher score indicates more signs of depression) has been widely used in studies of late-life depression and its psychometric properties are generally favourable [147]
- Kinesiophobia measured by the Norwegian version of the Fear Avoidance Beliefs Questionnaire - Physical Activity subscale (FABQ-PA) [148]. The FABQ-PA (range 0-24, higher score indicates higher levels of kinesiophobia) has been widely used and evidence supports its validity [148-150], including when used among Norwegian patients with back pain [151]
- Risk of poor disability outcome measured by the Norwegian version of the Keele STarT Back Screening Tool (SBST) [95]. The SBST is designed to screen primary care patients with low back pain for prognostic indicators that are relevant to initial decision-making.

The tool produces an overall score from 0-9 and a psychological subscale score from 0-5. Patients with an overall score between 0-3 are classified as low risk. Patients with a minimum overall score of 4 and a maximum subscale score of 3 are classified as medium risk. Patients with a minimum overall score of 4 and a subscale score of 4 or 5 are classified as high risk. The SBST has been recommended in guidelines to enable stratified care for patients with low back pain [43, 97]. The SBST was translated into Norwegian by Storheim and Grotle in 2012 and has been shown to be acceptably accurate in predicting persistent disabling back pain [95, 152-155]

- Expectations of recovery measured by the question *"What are your expectations for your back pain in 3 months?*" on a five-point scale and categorized into "recovered", "much better", or "no change or worse"
- Number of red flags (cancer, first episode of back pain, constant pain, unexplained weight loss, systematically unwell, fever, urinary retention or loss of bladder control, age ≥ 75 years, trauma as cause of back pain, osteoporosis, cortisone use, and severe morning stiffness) [33, 156]
- Total costs related to healthcare utilization during a period of 6 (BACE-N) or 12 (BACE-D) weeks prior to inclusion. Healthcare utilization prior to inclusion included: primary care consultations, use of back medication, and number of diagnostic examinations. Total cost of healthcare utilization per patient was estimated by multiplying frequency of use by unit costs collected from national pricelists (see section 3.8 for a detailed description of cost estimation)
- Physical workload at work measured by the Norwegian version [157] of the Physical Workload Questionnaire (PWQ) [158]. The PWQ (range 0-100, higher score indicates higher physical workload) consists of two subscales ("heavy physical workload" and "long-lasting postures and repetitive movements"), and evidence supports it validity among employees with musculoskeletal complaints [157, 158]
- Characteristics of psychosocial work environment (control of decisions, authorizing management, role conflicts, and fair leadership) measured by questions from the General Nordic Questionnaire for Psychological and Social Factors at Work (QPS Nordic) [159]. The QPS Nordic is a commonly used Nordic questionnaire for psychological and social factors in working life and evidence supports its reliability and validity [160]. Responses within the QPS Nordic are given in the form of a five-point Likert scale ("very seldom or never", "rather seldom", "sometimes", "rather often", and "very often or always")

Moreover, public registry data on long-term absenteeism (> 4 weeks) was obtained from NAV. Workers in Norway qualify for sickness benefits from NAV if they have been in paid work for the last 4 weeks before the sickness incident and if the occupational disability is documented by a sick leave certificate. In general, sickness benefit (100% of salary) can be received from the first day of reported sickness and for up to 1 year. If a person is still unable to work after 1 year, they may be entitled to disability benefits or work assessment allowance. Data on absenteeism obtained from NAV contains dates and grading of absenteeism in addition to the associated diagnostic code.

Rationale for selected variables in Papers I-IV

In Paper I, the SBST [95] was used as a screening tool to classify included participants into low, medium, or high risk of poor disability outcome.

In Paper II, based on previous scientific literature, the following abovementioned variables were included as potential modifiable prognostic factors: pain severity [64, 75, 102, 103, 109, 110], back-related disability [64, 75, 102, 103, 105, 109], health-related quality of life [102, 110], emotional well-being [75, 103, 105, 107, 108], kinesiophobia [75, 107], number of comorbidities [61, 110], radiating pain below the knee [75], and expectations of recovery.

Overall prognosis research (Paper I) and prognostic factor research (Paper II) may vary depending on context (time, place, healthcare setting) and the characteristics of the study population [12, 13]. Therefore, based on previous scientific literature, the following abovementioned variables were included as descriptive variables in Paper I: sex [64, 102, 161, 162], age [64, 102, 161, 162], educational level [105, 163], first healthcare provider [164], radiating pain below the knee [75], pain severity [64, 75, 102, 103, 109, 110], pain duration [103], pain history [109], back-related disability [64, 75, 102, 103, 105, 109], number of comorbidities [61, 108, 110], health-related quality of life [102, 110], emotional well-being [75, 103, 105, 107, 108], kinesiophobia [75, 107], number of red flags, and total costs related to healthcare utilization prior to inclusion. In Paper II, the following abovementioned variables were included as potential covariates: sex [64, 102, 161, 162], age [64, 102, 161, 162], educational level [105, 163], employment status, first healthcare provider [164], pain duration [103], pain history [109], and total costs related to healthcare utilization prior to inclusion.

In Paper III, based on previous scientific literature, the following abovementioned variables were included to investigate construct validity of the iPCQ: health-related quality of life [165, 166], pain severity [165, 167-170], physical workload at work [167, 171-173], and characteristics of psychosocial work environment [167, 168, 172-175].

In Paper IV, public registry data on long-term absenteeism (> 4 weeks) obtained from NAV was included to evaluate criterion validity of the iPCQ. Public registry data is often designated at the gold standard of absenteeism.

3.8 Analyses

An overview of analyses conducted in Papers I-IV is given in Table 5. All analyses in Papers I and II were outlined in a statistical analysis plan published a priori [126, 127]. All analyses were performed using the IBM SPSS version 24.0 (SPSS Inc., Chicago, IL) (Papers III and IV) and 26.0 (IBM Corporation, Armonk, NY, USA) (Papers I and II). Furthermore, in Paper III, the Vassarstats kappa was calculated using http://vassarstats.net/kappa.html. P-values below 0.05 were considered statistically significant. Paper II was considered to be explanatory, so no correction for multiple testing was performed within this paper. All statistical tests were two-sided. Papers I-IV were co-authored by a medical statistician. Papers I and II were co-authored by a health economist.

Table 5. Overview of analyses used in Papers I-IV

	Paper I	Paper II	Paper III	Paper IV
Study flow				
Frequency and percentage	X	X		
Mann-Whitney U test	X			
Pearson Chi-Square	X			
Fisher's exact test	X			
Descriptive statistics				
Frequency and percentage	X	X	X	X
Median and interquartile range or range	X	X	X	X
Mean and standard deviation	X	X	X	X
Data quality				
Frequency and percentage			X	
Missing data				
Frequency and percentage	X	X	X	X
Visually exploring	X	X		
Little's test	X	X		
Multiple imputation using regression estimation (baseline data)	X	X		
Content validity				
Relevance evaluating			X	
Comprehensiveness evaluating			X	
Comprehensibility evaluating			X	
Construct validity				
Confirmatory factor analysis			X	
Interitem correlation			X	
Hypothesis testing, Spearman's rho			X	
Criterion validity				
Percentage				X
Median and interquartile range				X
Intraclass correlation coefficient				X
Bland-Altman plot				X
Wilcoxon signed rank test				X
Spearman's rho				X
Reliability				
Intraclass correlation coefficient			X	
Cohen's unweighted kappa			X	
Healthcare utilization				
Frequency and percentage	X	X		
Median and interquartile range	X	X		
Kruskal-Wallis test incl. post hoc Mann-Whitney U tests with Bonferroni adjustment	X			
Cost estimation				
Frequency and percentage	X			
Mean with 95% CI using bias-corrected and accelerated bootstrapping	X	X		
Median with 95% CI using bias-corrected and accelerated bootstrapping	X	X		
Identification analysis				
Box-Tidwell transformation		X		
Univariable and multivariable binary logistic regression analyses		X		
Replication analysis				
Univariable and multivariable binary logistic regression analyses		X		
Sensitivity analyses				
Primary analyses conducted without adjustment for missing data	X	X		
Primary analyses conducted without outliers	X			

Study flow (Papers I and II)

Flow of participants through BACE-N (Papers I and II) and BACE-D (Paper II) were reported with flow charts according to the STROBE and the REMARK guidelines [176, 177]. Reasons for dropout were provided when known. Baseline differences between responders and non-responders at 12-month follow-up were evaluated. In Paper I, a Mann-Whitney U test was used for continuous variables, and a Pearson Chi-Square or Fisher's exact test (if < 5 cases in one cell) was used for categorical variables. In Paper II frequency and percentage were used.

Descriptive statistics (Papers I-IV)

Frequencies, medians, and means, and their respective variations (percentages, interquartile range or range, and standard deviations) were calculated to characterize the participants at baseline, and the iPCQ core items and index scores.

Data quality including missing data (Papers I-IV)

Floor and/or ceiling effects were assessed in Paper III and considered to be present if more than 15% of the participants reported the lowest or highest possible score [178]. The proportion of missing data was described with frequency and percentage in Papers I-IV.

In BACE-N (Papers I and II), the missing value pattern was visually explored, and missingness at random was assumed. Also, we found evidence against the hypothesis that values were not missing completely at random (Little's test, $p > 0.05$). Missing baseline data were handled by multiple imputation [179]. Five multiple imputation datasets with 10 iterations were created using regression estimation. We did not impute missing outcome values as the imputation model had poor predictive performance and caused a clear trend of values being overestimated. Instead, missing values on follow-up variables used to calculate the outcome cost scores were filled in with: 1) each patient's individual average of observed values for the variables: consultations with healthcare professionals and medication use, and 2) a value of zero costs for the variables: diagnostic examinations, hospitalization, rehabilitation stay, and back operations.

In BACE-D (Paper II), the missing value pattern was visually explored, and missingness at random was assumed. Missing values on follow-up variables used to estimate the outcome cost score were filled in with: 1) each patient's individual average of observed values for the variables: consultations with healthcare professionals and medication use, and 2) a value of zero costs for the variables: diagnostic examinations and back operations.

Healthcare utilization and cost estimation (Papers I and II)

Type and frequency of use of different healthcare resources were calculated for each of the follow-up periods in BACE-N (Papers I and II) and BACE-D (Paper II). Costs of healthcare utilization per patient were estimated by multiplying frequency of use by unit costs collected from national pricelists in Norway and the Netherlands (Table 6). Non-healthcare costs related to provision of healthcare (such as transportation) were not estimated. Costs related to back medication were estimated based on medication type (not exact medication name) and frequency of use converted to numbers of days with medication use (data on dosage were not available) (Table 7). All costs were presented in euros (€) for 2020 and estimated with both mean and median values with 95% CI, using bias-corrected and accelerated (BCa) bootstrapping. The BCa was conducted with a bootstrap sample size of 1000. Cost data are commonly skewed, so both mean and median values were presented to support result interpretation. Values in NOK were recalculated to euros using the exchange rate from February 2020 (€1 = NOK 10).

Table 6. Cost categories, units, unit price, all numbers in euros (€) for 2020

Cost categories	Unit	Norwegian unit price (€)	Dutch unit price (€)	Reference (source)
Primary care				
General practitioner	Per visit	43.1	36.0	The Norwegian Medical Association, estimated average iMTA costing tool [180]
Occupational physician	Per visit	-	36.0	iMTA costing tool [180]
Physiotherapist	Per visit	47.2	36.0	The Norwegian Physiotherapy Association, estimated average. iMTA costing tool [180]
Chiropractor	Per visit	55.0	36.0	Private price lists, estimated average
Manual therapist	Per visit	74.2	36.0	The Norwegian Physiotherapy Association, estimated average
Naprapath	Per visit	64.0	-	Private price lists, estimated average
Osteopath	Per visit	65.0	-	Private price lists, estimated average
Psychologist	Per visit	110.0	102.0	The Norwegian Psychological Association, estimated average. iMTA costing tool [180]
Other therapists	Per visit	75.0	-	Private price lists, estimated average
Back medication				
Paracetamol	Per daily defined dose	0.5	0.9	NoMA price list, estimated average. Medicijnkosten.nl, estimated average incl. pharmacy delivering costs [180]
NSAID	Per daily defined dose	1.2	0.4	NoMA price list, estimated average. Medicijnkosten.nl, estimated average incl. pharmacy delivering costs [180]
Muscle relaxant	Per daily defined dose	0.7	0.5	NoMA price list, estimated average. Medicijnkosten.nl, estimated average incl. pharmacy delivering costs [180]
Sleep medication	Per daily defined dose	0.2	-	NoMA price list, estimated average
Cortisone	Per daily defined dose	0.4	-	NoMA price list, estimated average
Opioid	Per daily defined dose	0.9	0.5	NoMA price list, estimated average. Medicijnkosten.nl, estimated average incl. pharmacy delivering costs [180]
Antidepressant	Per daily defined dose	-	0.3	Medicijnkosten.nl, estimated average incl. pharmacy delivering costs [180]
Anticonvulsant	Per daily defined dose	-	0.7	Medicijnkosten.nl, estimated average incl. pharmacy delivering costs [180]
Examinations				
Blood sample	Per examination	20.4	4.4	The Norwegian Medical Association, estimated average iMTA costing tool [180]
X-ray	Per examination	119.0	45.9	Unilabs price list, estimated average The National Health Authority
MRI	Per examination	269.0	233.0	Unilabs price list, estimated average iMTA costing tool [180]
CT	Per examination	189.0	151.0	Unilabs price list, estimated average iMTA costing tool [180]
Secondary care				
Medical specialist	Per visit	-	125.0	iMTA costing tool [180]
Back operation	Per operation	5220.0	5254.0	DRG2150. Different academic and non-academic hospital pricelists, estimated average
Hospitalization	Per day	1880.0	-	The Norwegian Directorate of Health, SAMDATA
Rehabilitation stay	Per day	315.0	-	UniCare pricelist, estimated average

iMTA indicates Institute for Medical Technology Assessment; NSAID, non-steroidal anti-inflammatory drug; NoMA, Norwegian Medicines Agency; €, euro. Cells marked with a dash (-) indicates that the unit price was not estimated.

Table 7. Assumptions made to estimate numbers of days with medication use*

Frequency	Estimated number of days with medication use last 3 months	
BACE-N		
Less frequently than every month	2	
Every month	6	(2 days a month)
Every week	26	(2 days a week)
Daily	91	
Several times a day	91	
BACE-D		
Less than once a week	6	(2 days a month)
Once or twice a week	19	(1.5 days a week)
Three to five times a week	52	(4 days a week)
Daily	91	

*Assumptions made in collaboration with health economist.

Healthcare utilization across patients with different risk profiles (Paper I)

Type and frequency of use of different healthcare resources were described for the one-year of follow-up for the following subgroups in BACE-N: 1) low, 2) medium, and 3) high risk of persistent disabling back pain according to the SBST. The Kruskal-Wallis test including post hoc Mann-Whitney U tests with Bonferroni adjustment were conducted to determine between-group differences with regards to: number of primary care consultations, number of patients using back medication, number of patients receiving imaging (X-ray, MRI, CT), and number of patients receiving secondary care (hospitalization, rehabilitation stays, back operations). The Bonferroni adjustment was applied by multiplying raw p-values by numbers of tests conducted (0.05x3).

High costs related to healthcare utilization (Paper II)

Costs related to healthcare utilization aggregated for one year of follow-up were dichotomized as high and low (in BACE-N and BACE-D). High costs related to healthcare utilization was defined as patients having costs in the top 25th percentile [75, 103].

Identification and replication analysis (Paper II)

First, univariable and multivariable binary logistic regression models were used to identify individual associations (crude and adjusted for selected covariates) between predefined modifiable prognostic factors and costs related to healthcare utilization (in BACE-N). The cost score was entered into the model as a dependent dichotomous variable (high costs defined as patients with costs in the top 25th percentile, yes/no). Linearity of continuous independent variables was examined using Box-Tidwell transformations [181]. Independent variables that demonstrated a non-linear relationship with the dependent variable were categorized before entering the model. Next, as described above, univariable and

multivariable binary logistic regression models were used to replicate findings from the identification analysis within the BACE-D material. Results were presented as crude and adjusted odds ratio (OR) with 95% CI. The decision on whether findings were replicated was based on the direction and size of the association and the size of the CI for each of the predefined modifiable prognostic factors [182].

Content validity (Paper III)

We investigated content validity by evaluating relevance, comprehensiveness, and comprehensibility, as recommended by COSMIN [124]. Two researchers and 1 clinician with no conflicts of interest were asked about the relevance and the comprehensiveness of the iPCQ (are all included items relevant for the construct of productivity cost and the target population, are all key elements of productivity costs included?). Ten patients with musculoskeletal disorders were asked open questions about the comprehensibility of the iPCQ (are the instructions, items, and responses understood as intended, are the items appropriately worded, and do the response options match the questions?).

Construct validity (Paper III)

We investigated construct validity by evaluating structural validity and internal consistency followed by hypothesis testing, as recommended by COSMIN [124]. To evaluate structural validity and confirm the underlying structure of the iPCQ, confirmatory factor analysis was conducted [183]. A computed factor loading expresses the magnitude or strength of an association between a given item and a factor (a component). The loading ranges from 0 to 1; the higher the value, the more an item is associated with a factor [183]. Based on the structure of the iPCQ, we expected that the core items would load onto 3 components: absenteeism (items no. 2, 3, 4, and 6), presenteeism (items no. 2, 3, 8, and 9), and unpaid work productivity costs (item no. 11 and 12).

To evaluate the internal consistency of the three components interitem correlation was assessed. For components to be considered sufficiently reliable, the interitem correlation should be > 0.4 [184].

The hypotheses (Table 8) included in the hypothesis testing were based on previous scientific literature [165-175]. Some inconsistency has been found in the literature. However, to the best of our knowledge, most available literature demonstrates a low correlation between productivity costs and health-related quality of life, pain severity, physical workload at work,

and psychosocial work environment. All hypotheses were investigated with Spearman's rho, as the scales were not normally distributed. Correlation coefficients < 0.3, between 0.3 and 0.6, and > 0.6 were considered low, moderate, and high, respectively [185]. According to recommendations, the iPCQ is considered to have adequate construct validity if at least 75% of the hypotheses are confirmed [178].

Table 8. Hypotheses for evaluating construct validity of the iPCQ in Paper I

1.	High productivity costs, assessed with 3 index scores of the iPCQ, will be negatively low correlated with low health-related quality of life [165, 166], assessed with the physical function and mental summary score of the SF36
2.	High productivity costs, assessed with the 3 index scores of the iPCQ, will be positively low correlated with high pain severity [165, 167-170], assessed with the NRS
3.	High productivity costs, assessed with the iPCQ index scores of absenteeism and presenteeism, will be positively low correlated with low physical workload at work [167, 171-173], assessed with the PWQ
4.	High productivity costs, assessed with the iPCQ index scores of absenteeism and presenteeism, will be positively low correlated with low psychosocial work environment [167, 168, 172-175], assessed with the QPS Nordic
5.	High productivity costs, assessed with iPCQ index score of productivity loss in unpaid work, will be positively low correlated with high number of pain locations [165, 167-170], assessed with the McGill pain drawing

iPCQ indicates the iMTA Productivity Cost Questionnaire; NRS, Numeric Rating Scale; PWQ, Physical Workload Questionnaire; QPS Nordic, General Nordic questionnaire for psychological and social factors at work; SF36, 36-Item Short Form Health Survey.

Criterion validity (Paper IV)

To evaluate criterion validity, the COSMIN group recommends evaluating the extent to which an instrument is an adequate reflection of a gold standard [125, 178]. Evaluating the criterion validity of iPCQ reported long-term absenteeism (> 4 weeks) can be done through comparison with public registry data. However, evaluation of the criterion validity of the remaining domains of the iPCQ poses significant challenges due to the lack of a gold standard or objective measures [186]. To evaluate the criterion validity of iPCQ reported long-term absenteeism (> 4 weeks), data on long-term absenteeism (reported by the iPCQ and the public registry) were used to generate the following variables:

a) Occurrence of long-term absenteeism. Defined as a continuous period of more than 4 weeks of absenteeism recorded retrospectively from baseline (yes/no).
b) Duration of long-term absenteeism. Defined as the duration of a continuous period of absenteeism from baseline to maximum 12 months retrospectively. The duration of long-term absenteeism was operationalized in two different ways: 1) by calculating the number of calendar days from start date until end date of sick leave (defined as the date the iPCQ was completed) (duration), and 2) by adjusting for grading of absenteeism, summarizing number of days with part-time sick leave to number of days with complete sick leave

(adjusted duration) (e.g., 10 days with 50% sick leave equals absenteeism duration and adjusted duration of 10 and 5 days, respectively).

To compare the occurrence of long-term absenteeism, participants were classified according to whether a continuous period of long-term absenteeism had been recorded by the iPCQ (yes/ no) and the public registry (yes/no). The overall agreement between the two methods was expressed as: (number of identical/total answers) × 100. To compare the duration and adjusted duration of long-term absenteeism, we computed intraclass correlation coefficient (ICC) using two-way random average agreement. The acceptable level of ICC was set to > 0.70 [178]. Furthermore, to illustrate the relationship between the two methods, we depicted their $\text{differences}_{(\text{iPCQ-registry})}$ and averages using Bland-Altman plots. Further, the $\text{differences}_{(\text{iPCQ-registry})}$ were described with medians and interquartile ranges and analysed with the Wilcoxon signed rank test. To test whether differences between the two methods were associated with the length of sick leave, as recorded by the public registry, stratified analyses for the following categories of absenteeism length were performed: ≤ 3 months, > 3 months to ≤ 6 months, and > 6 months. In addition, Spearman's rho was used to assess the correlation between the $\text{differences}_{(\text{iPCQ-registry})}$ and the length of sick leave. Correlation coefficients < 0.3, between 0.3 and 0.6, and > 0.6 were considered low, moderate, and high, respectively [185].

Reliability (Paper III)

We evaluated reliability of the iPCQ by assessing test-retest reliability. Continuous variables (items no. 2, 3, 4 second part, 6, 8, 9, 11, and 12) and the 3 index scores of the iPCQ were assessed with the intraclass correlation coefficient (ICC) using 2-way random, average agreement. Dichotomous variables (items no. 4 first part, 5, 7, and 10) were assessed with Cohen's unweighted kappa. The acceptable level of ICC was set to > 0.70 [178]. Kappa values were categorized according to Altman: poor (0 to 0.2), fair (0.21 to 0.40), moderate (0.41 to 0.60), good (0.61 to 0.80), and very good (0.81 to 1.00) [187].

Sensitivity analyses (Papers I, II, and IV)

To assess the credibility of total cost calculations related to the primary analyses in Paper I, two sensitivity analyses were performed: 1) complete case analysis without adjustment for missing data, and 2) without outliers. Outliers were identified with simple scatterplots by visual inspection and defined as patients with remarkably high total costs at each time period: 5 patients with costs ≥ €2433 at 0-3 months, 5 patients with costs ≥ €6025 at > 3-6 months, 8 patients with costs ≥ €3518 at >9-12 months, and 11 patients with costs ≥ €8004 at 0-12

months. To assess the credibility of the identification analysis and possible bias introduced by the multiple imputation procedure in Paper II, the univariable and multivariable binary logistic regression analyses were performed on complete case data (in BACE-N).

Sample Size (Papers I-IV)

Papers I, II, and IV contain secondary analyses embedded in BACE-N, BACE-D, and the MI-NAV project. Details on sample size calculations related to the primary aims of these cohorts are provided in the study protocols [34, 35, 123]. In Paper I, we considered a sample size of 450 participants sufficient to describe healthcare utilization and estimate associated costs [188]. In Paper II, we used number of events per variable (EPV) [189-193] and the rule of thumb of "10 events per 1 analysed variable" [194-197] to determine statistical power. With a minimum sample size of 450 participants in BACE-N and BACE-D, a minimum of 112 participants were anticipated to be in the top 25th percentile of costs and defined as having high costs related to healthcare utilization (yes/no) (events). An EPV of 10 would allow a maximum of 11 prognostic variables to be included in the logistic regression models. We used quality criteria recommended by Mokkink et al. [125] to determine statistical power for Paper IV; a sample size of 30 to 50 participants in the smallest group is considered adequate when investigating criterion validity.

In Paper III, we planned to recruit 100 participants based on the quality criteria recommended by Terwee et al. [178] and Nunnally et al. [184]. These criteria suggest a minimum of 100 participants for assessing internal consistency, at least 50 participants for assessing reproducibility and floor or ceiling effects [178], and at least 10 participants for each item included in the factor analysis [184].

SUMMARY OF RESULTS

The main results from Papers I-IV are presented below. Participants included in the four papers are presented together in section 4.1, while the results from the papers are presented separately in sections 4.2 to 4.5. Detailed results including tables and figures are found in the enclosed papers.

4.1 Participants Papers I-IV

Papers I and II included 452 patients with back pain aged ≥ 55 years (BACE-N). In addition, Paper II included 675 patients with back pain aged > 55 years (BACE-D). Paper III included 115 patients with musculoskeletal disorders (87% with back or neck pain) aged 21 to 65 years. Finally, Paper IV included 144 people on long-term sick leave due to musculoskeletal disorders (19% with back or neck pain) aged 24 to 67 years. The characteristics and clinical status of included participants at baseline in Papers I-II and Papers III-IV are presented in Table 9 and 10, respectively.

Flow of participants through Papers I and II is shown in Figure 1. Fourteen patients (3%) in BACE-N and 22 patients (3%) in BACE-D were dropouts at 12-month follow-up and were not included in the analyses. In BACE-N, there was a larger proportion of females (55 vs. 42%) among responders as compared to non-responders. In BACE-D, there was a larger proportion of people not in paid work (26 vs. 38%) and people with short pain duration < 6 weeks (56 vs. 39%) among responders as compared to non-responders. There were no other differences between responders and non-responders in the two cohorts. The BACE-N and BACE-D samples were largely comparable, though BACE-N had a larger proportion of people with high education level (44 vs. 17%), people in paid work (47 vs. 27%), and people with pain duration < 6 weeks (67 vs. 54%).

Table 9. Characteristics and clinical status at baseline of participants included in Papers I and II

	Paper I	Paper II	
	BACE-N (n = 452)	**BACE-N (n = 452)**	**BACE-D (n = 675)**
Female	235 (52)	235 (52)	401 (59)
Age in years	66 (59-72)	66 (59-72)	65 (60-71)
Education level high	188 (44)	188 (44)	114 (17)
Ethnicity Norwegian (BACE-N) or Dutch (BACE-D)	430 (95)	430 (95)	637 (96)
Employment status currently in paid work	211 (47)	211 (47)	177 (27)
Productivity loss (iPCQ)			
Absenteeism last 4 weeks	75 (17)	-	-
Presenteeism last 4 weeks	132 (32)	-	-
Productivity loss unpaid work last 4 weeks	135 (35)	-	-
First healthcare provider			
General practitioner	127 (28)	127 (28)	675 (100)
Physiotherapist	130 (29)	130 (29)	0 (0)
Chiropractor	195 (43)	195 (43)	0 (0)
Pain location			
Thoracic	56 (13)	56 (13)	154 (26)
Lumbar/sacral	406 (92)	406 (92)	561 (93)
Radiating pain below the knee	141 (31)	141 (31)	205 (31)
Pain severity average last week (NRS, 0-10)	5 (4-7)	5 (4-7)	5 (3-7)
Pain duration			
< 6 weeks	252 (67)	252 (67)	323 (54)
6 weeks to 3 months	49 (13)	49 (13)	116 (20)
> 3 months	75 (20)	75 (20)	156 (26)
Previous episodes of back pain	400 (95)	400 (95)	579 (86)
Disability (RMDQ, 0-24)	9 (4-13)	9 (4-13)	10 (5-14)
Comorbidity (SCQ, 0-15)	1 (1-2)	1 (1-2)	2 (1-3)
Health-related QOL (SF36, 0-100)			
Physical component	42 (36-47)	42 (36-47)	43 (37-50)
Mental component	55 (48-60)	55 (48-60)	52 (43-57)
Emotional well-being (CES-D, 0-60)	8 (3-13)	8 (3-13)	9 (4-14)
Kinesiophobia (FABQ-PA, 0-24)	9 (5-13)	9 (5-13)	14 (10-17)
Expectations of recovery within 3 months			
Fully recovered	-	111 (26)	113 (17)
Much better	-	217 (50)	178 (27)
No change or worse	-	105 (24)	367 (56)
Healthcare utilization prior to inclusion			
Primary care consultation last 6 (BACE-N) or 12 (BACE-D) weeks			
General practitioner	78 (18)	78 (18)	609 (91)
Occupational physician	-	-	13 (2)
Physiotherapist, chiropractor, or manual therapist	295 (68)	295 (68)	299 (45)
Psychologist	2 (0.5)	2 (0.5)	5 (1)
Other therapists	21 (5)	21 (5)	-
Use of back medication	165 (40)	165 (40)	484 (73)
Diagnostic examination last 6 (BACE-N) or 3 (BACE-D) months			
Blood sample	12 (3)	12 (3)	92 (14)
X-ray	23 (5)	23 (5)	155 (23)
MRI/CT scan	49 (11)	49 (11)	30 (5)
Previous hospitalization	48 (11)	48 (11)	-
Previous rehabilitation stay	17 (4)	17 (4)	-
Medical specialist consultation	-	-	46 (7)

CES-D indicates the Center for Epidemiologic Studies Depression Scale; FABQ-PA, The Fear Avoidance Beliefs Questionnaire, physical activity subscale; iPCQ, The iMTA Productivity Cost Questionnaire; NRS, Numeric Rating Scale; RMDQ, The Roland Morris Disability Questionnaire; SCQ, The Self-Administered Comorbidity Questionnaire; SF-36, 36-Item Short-Form Health Survey. The presented characteristics are based on complete case data. All values are presented by number (valid percentage of total) or median (IQR). Cells marked with a dash (-) indicates that the variable was not measured. The original article for Paper I presented pooled estimates based on multiple imputation procedures. For the purpose of this table, all values were based on complete case data to give similar values for both papers.

Table 10. Characteristics and clinical status at baseline of participants included in Papers III and IV

	Paper III		Paper IV
	Validity study (n = 115)	**Test-retest study (n = 62)**	**(n = 144)**
Female	79 (69)	39 (63)	85 (59)
Age in years	45 (21-65)	46 (28-65)	49 (24-67)
Education level high	67 (58)	35 (57)	71 (49)
Mother tongue Norwegian	100 (87)	53 (86)	128 (89)
Employed or self-employed (paid job)	104 (90)	55 (89)	144 (100)
Weekly work hours	38 (8-52)	38 (8-52)	38 (4-60)
Weekly workdays	5 (2-7)	5 (2-7)	5 (2-7)
Absenteeism last 4 weeks (iPCQ)	79 (69)	40 (65)	144 (100)
Type of absenteeism			
Partial sick leave	-	-	17 (12)
Complete sick leave	-	-	48 (33)
Partial and complete sick leave	-	-	79 (55)
Presenteeism last 4 weeks (iPCQ)	32 (28)	23 (37)	68 (47)
Productivity loss unpaid work last 4 weeks (iPCQ)	52 (45)	29 (47)	75 (52)
Pain severity last 1 (Paper III) or 2 weeks (Paper IV) (NRS, 0-10)	5 (1-9)	5 (1-9)	5 (1-9)
Pain location			
Lower limbs	70 (61)	36 (58)	22 (15)
Back and neck	100 (87)	54 (87)	28 (19)
Upper limbs	55 (48)	24 (39)	41 (29)
≥ Two pain areas	45 (39)	23 (37)	30 (21)
Others	-	-	23 (16)
Health-related quality-of-life (SF36, 0-100)			
Mental health	70 (10-100)	75 (10-95)	-
Physical function	75 (30-100)	70 (30-95)	-
Physical workload (PWQ, 0-100)			
Heavy physical workload	21 (0-86)	26 (0-86)	-
Long-lasting posture and repetitive movement	50 (0-100)	50 (0-100)	-
QPS Nordic (1-5)			
Control of decisions	3 (1-5)	3 (1-5)	-
Authorizing management	3 (1-5)	4 (1-5)	-
Role conflict	2 (1-5)	2 (1-5)	-
Fair leadership	4 (1-5)	4 (1-5)	-

iPCQ indicates the iMTA Productivity Cost Questionnaire; NRS, Numeric Rating Scale; PWQ, Physical Workload Questionnaire; QPS Nordic, General Nordic questionnaire for psychological and social factors at work; SF36, 36-Item Short Form Health Survey. The presented characteristics are based on complete case data. All values are presented by number (valid percentage of total) or median (range). Cells marked with a dash (-) indicates that the variable was not measured. The original article for Paper III presented mean (SD) value for age and separate values for back and neck pain location. The original article for Paper IV presented median (IQR) values for weekly workhours and workdays and mean (SD) value for pain severity. For the purpose of this table, values were recalculated to give comparable values between the two papers.

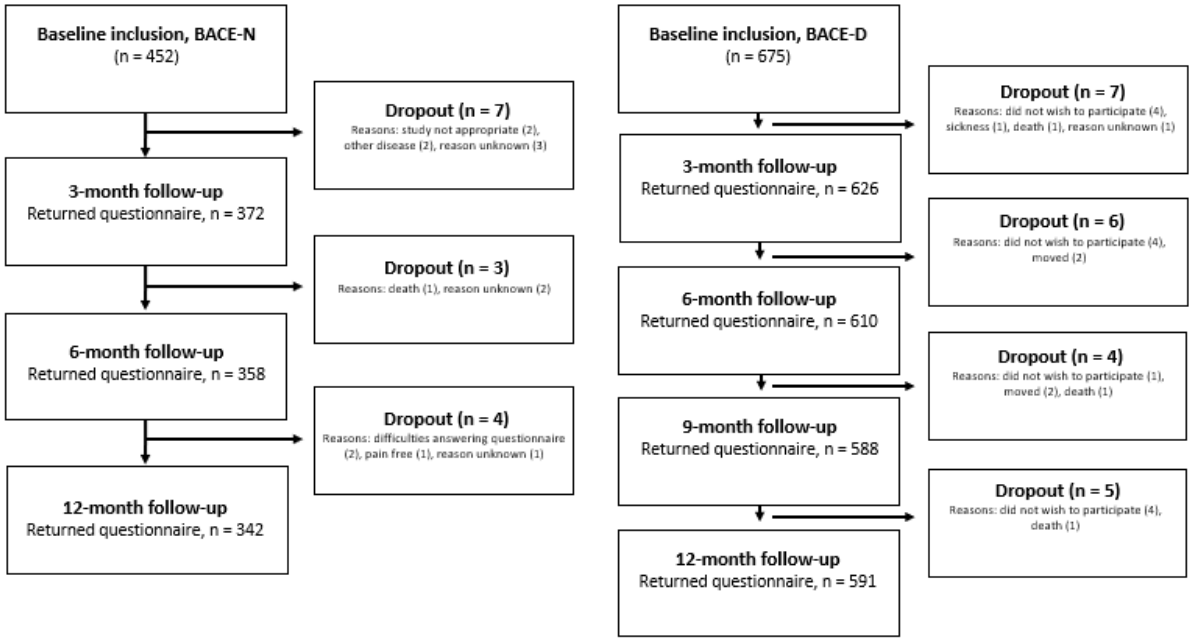

Fig 1. Participant flow chart BACE-N and BACE-D

4.2 Paper I

Healthcare utilization and related costs among older people seeking primary care due to back pain

Almost all included patients (87%) had costs related to healthcare utilization during the one-year of follow-up. Mean (BCa 95% CI) and median (BCa 95% CI) total cost per patient was €825 (682-976) and €364 (307-440), respectively. Patients within the top 25th percentile accounted for 77% of total costs within the sample. The largest cost category was primary care consultations (56% of total costs). The remaining cost categories were hospitalization (16%), back operations (11%), examinations (8%), back medication (6%), and rehabilitation stays (3%). Imaging rate was 34% during the one-year of follow-up, including the time period from baseline to 6 months retrospectively. A total of 34-45% of patients used medication, and the most commonly used medication was paracetamol (27-35% of patients).

Patients with medium and high risk of persistent disabling back pain had a significantly higher degree of healthcare utilization compared to patients with low risk (Kruskal-Wallis test, post hoc Mann-Whitney U tests with Bonferroni adjustment, ($p < 0.030$)). No statistically significant differences were revealed between medium- and high-risk patients.

4.3 Paper II

Modifiable prognostic factors of high costs related to healthcare utilization among older people seeking primary care due to back pain

In BACE-N and BACE-D, a total of 110 and 163 patients were defined as having high costs (≥ €789 and ≥ €664) during the one year of follow-up, respectively. Figure 2 shows adjusted OR estimates with 95% CI for the association between each of the modifiable prognostic factors and being in the high costs group. In both the identification and replication analysis, after adjustment for selected covariates, factors associated with increased odds of being in the high costs group were a higher degree of pain severity, disability, and depression, and a lower degree of physical health-related quality of life. No associations were found between being in the high costs group and the degree of kinesiophobia or expectations of recovery. There were inconsistent results across the two cohorts with regards to the impact of comorbidity, radiating pain below the knee, and mental health-related quality of life on high costs related to healthcare utilization.

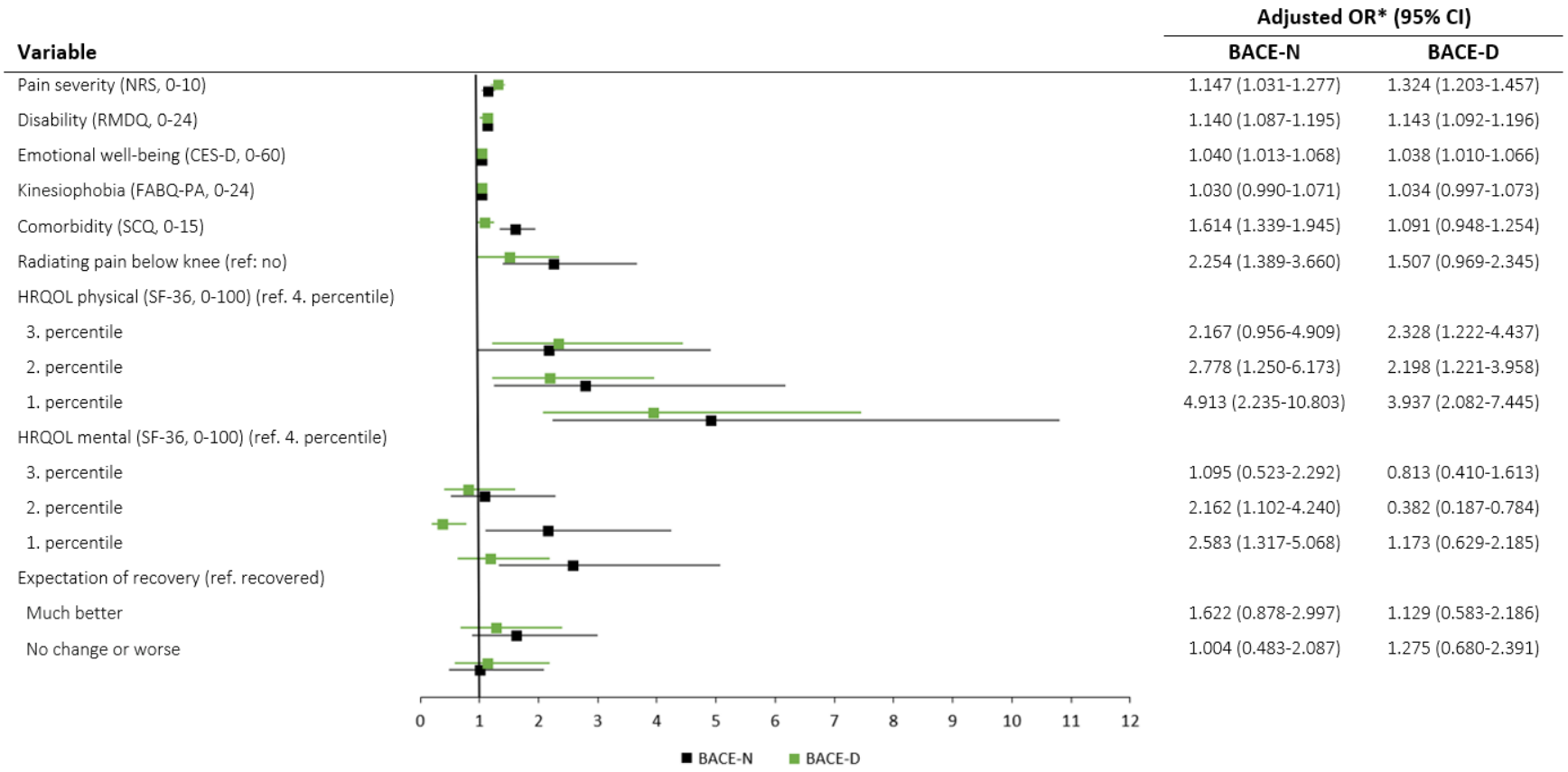

Fig 2. Forest plot summary of binary logistic regression analyses for individual associations between modifiable prognostic factors and high costs related to healthcare utilization. Adjusted OR (boxes) and the corresponding 95% CI (lines) are shown in black and green for the BACE-N and BACE-D sample, respectively. CES-D indicates The Center for Epidemiologic Studies Depression Scale; CI, confidence interval; FABQ-PA, The Fear Avoidance Beliefs Questionnaire, physical activity subscale; HRQOL, health-related quality of life; NRS, Numeric Rating Scale; OR, odds ratio; RMDQ, The Roland Morris Disability Questionnaire; SCQ, The Self-Administered Comorbidity Questionnaire; SF-36, The Short-Form Health Survey 36-item. *Adjusted by sex, age, education level, employment status, pain duration, pain history, first healthcare provider and costs related to healthcare utilization prior to inclusion.

4.4 Paper III

Measurement properties of the iMTA Productivity Costs Questionnaire

The proportion of missing data was < 10% for all items and there were no floor or ceiling effects. Content validity was assessed to be sufficient as all the included items were relevant and covered all domains of productivity costs, except compensation mechanisms and part-time sick leave. Moreover, the iPCQ was understood as intended by the 10 patients evaluating comprehensibility. Construct validity was confirmed. The confirmatory factor analysis revealed a three-component solution accounting for 82% of the total variance in the data. Moreover, as expected, the analysis showed that core items no. 2, 3, 4, and 6 loaded on the first component (absenteeism), that core items 2, 3, 8, and 9 loaded on the second component (presenteeism), and that core items 11 and 12 loaded on the third component (productivity loss unpaid work). The internal consistency and the level of inter-item correlation were acceptable for all components of the iPCQ with values of 0.46, 0.42, and 0.62 for absenteeism, presenteeism, and productivity costs from unpaid work, respectively. A total of 91% of our hypotheses were confirmed. Test-retest reliability was above the recommended minimum standard for all core items and the index scores (ICC ≥ 0.88), except for item no. 8 (number of workdays with disability, ICC = 0.34). Kappa values of the four dichotomous items of the iPCQ ranged from 0.62 to 0.84.

4.5 Paper IV

Criterion validity of the iMTA Productivity Cost Questionnaire

Occurrence of self-reported long-term absenteeism assessed by the iPCQ was identical to public registry data with 100% agreement. Duration (no. of days with absenteeism) of long-term absenteeism assessed by the iPCQ correlated highly with public registry data, with an ICC (95%CI) of 0.93 (0.91-0.95). Furthermore, descriptive statistics displayed a median (IQR) difference$_{(iPCQ\text{-}registry)}$ of 0 (0-0) days between the two methods. Adjusted duration (no. of days with complete absenteeism) of long-term absenteeism assessed by the iPCQ correlated acceptably with public registry data, with an ICC (95%CI) of 0.75 (0.48-0.86). Compared to the public registry data, the participants overestimated the numbers of days with long-term absenteeism with median (IQR) 17 (0-49) days, and a statistically significant difference between the two methods was revealed (Wilcoxon signed-rank test, $p < 0.001$).

DISCUSSION

5.1 Methodological considerations

In this section, methodological aspects of Papers I-IV are discussed with respect to internal and external validity. Formal checklists for quality analysis of overall prognosis studies and COI studies are lacking. Thus, domains from the Quality In Prognosis Studies (QUIPS) tool [198] and the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement [199] were used when evaluating the internal and external validity of Paper I. The QUIPS tool [198] was used as the main framework for Paper II. The COSMIN Bias Checklist [200] was used for Papers III and IV.

5.1.1 Internal validity

Papers I and II

Based on the QUIPS tool [198] and the CHEERS statement [199], the following aspects of internal validity were evaluated for Papers I and II: study attrition, outcome measurement, prognostic factor measurement, subgroup measurement, study confounding, and statistical analysis and reporting.

A low dropout rate (3%) and relatively high follow-up response rate (76-93%) indicated a low risk of attrition bias in Papers I and II [2]. Although there is no agreed upon standard for an acceptable response rate, a rate of 60% has been used as a rule of thumb for an acceptable threshold in surveys [201], and a rate greater than 85% is often considered high [2]. However, some possibly important differences between responders and non-responders were seen. In BACE-N, there was a larger proportion of females (55 vs. 42%) among the responders as compared to non-responders. In BACE-D, there was a larger proportion of people not in paid work (26 vs. 38%) and people with shorter pain duration (56 vs. 39%) among the responders as compared to non-responders. Previous studies have shown that females are more likely to seek care for back pain as compared to males [64, 102, 109], as are people with persistent pain [103]. Hence, it is likely that the amount of healthcare utilization presented in BACE-N is somewhat overestimated, whereas the opposite may be true in BACE-D.

The risk of outcome measurement bias is considered to be low to moderate in Paper I, based upon the following arguments: 1) it is assumed that all relevant cost components were included and that methods of measurement were adequate, valid, and reliable, 2) costs unrelated to back pain were omitted, 3) all unit costs were valued appropriately in

collaboration with a health economist and data sources were described, 4) method and setting of measurement were the same for all participants if we disregard the fact that some questionnaires were completed on paper, 5) the rates of missing data on healthcare utilization variables were in the lower layer (18-26%), and 6) missingness at random was assumed. However, the fact that we had to manually replace missing values on variables used to calculate costs may have introduced some risk of bias. It is well-known that healthcare utilization is prone to missing data [202-204], and that missing data should be replaced in order to make use of all reported data [202, 203]. Unfortunately, due to poor predictive performance, model-based multiple imputation could not be used on the follow-up data. We therefore chose a frequently used, though not optimal, method for replacing missing values [204], and have been transparent in our reporting (see section 3.8 for a detailed description of the method used). The lack of data covering primary care consultations and medication use between 6 and 9 months in BACE-N is also expected to cause a risk of bias and might have led to an underestimation of total healthcare utilization and related costs in BACE-N. Still, reported healthcare utilization of these resources was highest at the beginning of the follow-up period (0-3 months as compared to > 3-6 months and > 9-12 months), and it is therefore only expected to have a minor impact on the results. Finally, we, like others [20], used self-reported data on healthcare utilization. Self-reported data may be affected by recall bias, and it has been shown that self-reports tend to underestimate the true value of healthcare utilization [205, 206]. This might be especially relevant among older people [207, 208]. Therefore, once again, we expect to have somewhat underestimated total healthcare utilization and related costs in Paper I. In Paper II, the following elements indicate a low risk of bias regarding the outcome measurement: 1) a clear definition of high costs related to healthcare utilization was provided, 2) our measurement of healthcare utilization and costs calculation is assumed to be adequate, as discussed above, 3) the rate of missing data on variables used to calculate the outcome variable was in the lower layer (7-26%), and 4) missingness at random was assumed. However, as mentioned above, the fact that we had to manually replace missing values on variables used to calculate the outcome score may have introduced some risk of bias. Nevertheless, all participants with costs in the top 25th percentile in the complete case sample were also participants with costs in the top 25th percentile after missing values were replaced. Furthermore, it might have introduced some risk of bias that costs related to hospitalization and rehabilitation stays were not measured in BACE-D. Thus, the risk of misclassification bias related to whether participants were classified as having high or low costs might be present in BACE-D. However, if costs related to hospitalization and

rehabilitation stays were omitted from the cost calculations in BACE-N, only 4 participants (<1%) switched cost group. Therefore, we argue that the dichotomization of the outcome variable is trustworthy. Likewise, we consider the impact of recall bias to be minor in Paper II since the outcome variable was dichotomized into high or low costs.

The risk of bias in the prognostic factor measurement is assumed to be low to moderate in Paper II. Firstly, each of the prognostic factors was clearly defined and measured with commonly used and well-validated questionnaires (except for the factors "radiating pain below the knee" and "expectation of recovery", which were measured with one-item questions). However, none of the measurement tools have been formally validated for Norwegian older people with back pain, which may have introduced some risk of bias. Based on normative data from the general population in Norway showing that data-completeness of the SF-36 declined with increasing age, it has been emphasized that caution should be exercised when using this tool among people aged ≥ 70 years [209]. Furthermore, our measurement of comorbidity (the SCQ) might not have captured all relevant comorbidities for an older population with back pain. For example, 7% of the BACE-N sample reported osteoporosis in the open-ended SCQ items, making osteoporosis the fifth most common comorbidity in BACE-N [210]. Secondly, the rate of missing data on the prognostic factor variables was low (0-13%) and an appropriate method of imputation (multiple imputation) [179] was used for missing data within the identification sample (BACE-N).

The risk of misclassification bias related to the SBST subgroup classification is considered to be low in Paper I. Firstly, the SBST has been validated for older patients with back pain and has been found to provide an acceptable accuracy for distinguishing between patients with low and high probability of persistent disabling back pain [155]. Secondly, the rate of missing data on the SBST was low (7%) and an appropriate method of imputation (multiple imputation) [179] was used for missing data on this tool. A potential weakness, however, is that up to 82% of people seeking primary care due to back pain are expected to change SBST risk profile within 4 weeks from intake [211]. Thus, it could be argued that repeated risk assessment with the SBST would have increased grouping credibility. However, in BACE-N there was a large predominance of low-risk patients, of which only 11% were expected to change risk profile [211]. Importantly, we aimed to describe healthcare utilization across patients with different risk profiles at baseline, regardless of whether they changed group affiliation over time.

Overall, the risk of cofounding bias is considered to be low in Paper II. A clear definition of each covariate was provided, and methods of measurement are assumed to be adequate. The rate of missing data on covariates was low (0-17%), and an appropriate method of imputation [179] was used for missing data within the identification sample (BACE-N). Nevertheless, the fact that we could not adjust for some potentially important covariates of healthcare utilization, such as the patient's disposition to access and pay for healthcare services and health insurance status, might have introduced some risk of bias [101]. However, these factors are likely of less importance in countries such as Norway and the Netherlands, where health services are widely available and largely covered by the public sector.

The risk of analysis and reporting bias is assumed to be low in Papers I and II. Both papers were thoroughly pre-planned together with a group of skilled back pain researchers and a statistician, including a statistical analysis plan published a priori [126, 127]. Furthermore, Paper I adheres to the STROBE [176] and the CHEERS [199] statements reporting guidelines, while Paper II adheres to the REMARK reporting guidelines [177]. However, our result in Paper I showing no difference in healthcare utilization between medium- and high-risk patients should be interpreted with caution due to the small sample size within the high-risk subgroup, and thus a risk of low statistical power. In Paper II, our categorization of the two SF-36 variables based on percentiles may have introduced some risk of analysis bias [12, 212]. This categorization is not recommended according to the PROGRESS framework [12]. Nevertheless, these variables demonstrated a non-linear relationship with the dependent variable and could not be entered into the model as such. Therefore, after consulting a statistician, and for practical reasons due to the replication analysis, we categorized these variables based on percentiles.

Papers III and IV

By following the COSMIN Bias Checklist [200], the overall methodological quality of Papers III and IV can be considered inadequate to very good, depending upon the different aspects of measurement quality assessed. Our assessment of structural and criterion validity and the hypothesis testing can be considered very good, whereas the assessment of content validity, internal consistency, and reliability are weaker. Content validity was assessed by adhering to the COSMIN guidelines at the time the validation study was conducted (2015) [124]. Two researchers and one clinician, with no conflicts of interest, provided input on relevance and comprehensiveness. Furthermore, 10 patients were asked about comprehensibility. In 2017, the COSMIN guidelines for assessing content validity of patient-reported outcome measures

were updated [5] recommending that patients should also evaluate relevance and comprehensiveness, professionals from all relevant disciplines should evaluate relevance and comprehensiveness, and interviews should be conducted by skilled interviewers, as well as recorded and transcribed. Since these aspects are lacking in our content validity assessment, our findings ought to be reassessed. We used inter-item correlation estimates to assess the internal consistency of the three iPCQ components, which, according to Pallant [183], is the most appropriate method to use when each component has < 10 items. COSMIN, however, recommends the use of Cronbach's alpha, which was not included in Paper III. As for the reliability assessment, we did not provide evidence that the participants were stable within the interim period. However, at re-test ≥ 92% of the participants reported no change in working situation, and ≥ 86% reported no or only small changes in symptoms. Regrettably, this information was omitted from the published article where it should undoubtedly have been included.

5.1.2 External validity

Papers I and II

Based on inclusion and exclusion criteria, and the method of recruitment, we can assume that the results from Papers I and II are applicable to Norwegian-speaking people aged ≥ 55 years seeking Norwegian primary care with a new episode of back pain. In addition, for Paper II, we can also assume that the results are applicable to Dutch-speaking people aged > 55 years seeking a GP in the Netherlands with a new episode of back pain. The major weakness concerning the generalizability of Papers I and II is that we lack data on eligible participants that declined to participate or were not invited to join BACE-N and BACE-D during the data collection period. Unfortunately, gathering this data was not possible due to limited resources and recruitment from a broad network of clinicians. In order to compensate for this limitation and explore the representativeness of the BACE-N sample, its key sociodemographic variables have been compared with those of the Norwegian study on life course, aging and generation (NORLAG), a large population study on older people in Norway [213, 214]. A subsample of the NORLAG participants (NORLAG MSK) was used, representing people aged ≥ 55 years with musculoskeletal complaints. The characteristics of the two samples were largely comparable, though BACE-N had more men (48 vs. 36%) and more participants with higher education level (44 vs. 29%) as compared to the NORLAG MSK [214]. Furthermore, the BACE-N sample is largely comparable to younger Norwegian back pain cohorts [215, 216] as well as the BACE-D sample [34].

Finally, an issue that can be discussed is the definition of an older person in Papers I and II. Using 55 years as an age cut-off point might have affected whether the BACE-N and BACE-D samples represent an older population. Commonly, older people are defined as those aged 60 or 65 years or older [217]. In BACE-N and BACE-D, 74% and 80% of patients were ≥ 60 years at baseline, respectively, and 58% and 52% of patients were ≥ 65 years at baseline. The age cut-off point of ≥ 55 years in BACE-N and BACE-D was determined based on the standardized methodology of the BACE consortium [34], as this would allow for comparisons across different countries. Within the BACE consortium, the decision regarding the age cut-off point was based on the age cut-off point (of ≥ 55 years) used in a large population cohort study of older people in the Netherlands (The Rotterdam Study) [218].

It is uncertain to what extent the results from Papers I and II apply to people outside of the target populations. Due to differences in primary care organizations between countries, caution is warranted when generalizing the results to other countries and healthcare systems. However, the external validity of Paper II is strengthened by our findings being replicated in another country.

Papers III and IV

For Paper III, we assume that the results are applicable to Norwegian-speaking adult patients with musculoskeletal disorders who receive treatment at an outpatient rehabilitation clinic in Norway. However, due to limited resources, we lack information on eligible participants that declined to participate or were not invited to join the study. Thus, the risk of selection bias is present in Paper III. To compensate for this limitation, we instructed the recruiting clinicians to invite all eligible patients to participate within the recruiting period [219]. Nevertheless, it is not possible to achieve the ideal circumstances in which everyone could be asked and have an equal chance of joining the study [219]. Further, eligible patients could choose whether they wanted to participate, and it may be that the people who agreed differed from those who did not [219].

For Paper IV, results are assumed to be applicable to adult people on long-term sick leave due to musculoskeletal disorders in Norway who have the ability to read and write in Norwegian or English. The study sample in Paper IV is comparable to the total MI-NAV sample in terms of age, gender, and education level [220, 221]. Representativeness of the total MI-NAV sample has been evaluated previously [220], and the demographic and socio-economic characteristics of participants resembled non-participants (n = 168.137), broadly

confirming that the total MI-NAV sample was representative of the target population [220]. The risk of selection bias is therefore assumed to be low in Paper IV.

It is uncertain to what extent the results from Papers III and IV apply to people outside the target populations. However, it is likely that results can be cautiously generalized to other settings, such as primary care, and to other populations.

5.2 Main results compared with other studies

In this section, the main results of Papers I-IV are discussed in the context of previous research. In addition, the main results of Paper I are discussed in the context of clinical guidelines [43, 92, 93].

Papers I and II

Direct comparability of Papers I and II with other studies is limited as, to the best of my knowledge, no similar studies have been conducted among a sample of exclusively older people with back pain or within the Norwegian healthcare system [20]. Furthermore, with regards to Paper I, there is great heterogeneity in the methods used among COI studies on back pain, as illustrated in Table 1, thus direct comparability between studies is not feasible [19, 20]. Nevertheless, several of our findings are generally in accordance with previous research on primarily middle-aged patients with back pain [20, 64, 75, 84, 100, 102, 103, 105, 107, 109, 113], as well as patients with musculoskeletal disorders [110-112].

In Paper I, we estimated a one-year mean total direct cost per patient of €825, in 2020 euros. Patients within the top 25th percentile accounted for 77% of all costs, and the largest cost category was primary care consultations (56% of total costs). An estimated one-year mean total cost per patient of €825 is more or less in line with previous research, yet in the lower layer of what has been reported (see Table 1) [20]. Previously, the one-year mean total direct cost related to back pain per patient has been estimated in 2020 euros to range from €761 in Germany [64] to €13.783 in Japan [68]. However, the majority of studies recruiting participants from primary care estimated costs ranging from €1.000 to €2.000, in 2020 euros [20, 69, 70, 75]. Furthermore, it is in accordance with previous research that the majority of healthcare utilization and related costs stem from a relatively small group of patients with back pain [84, 100, 113], and that primary care consultations are frequently used and constitute a large cost category among these patients [19, 20, 56, 64, 69, 70, 78, 81, 84].

We found an imaging rate of 34% during the one year of follow-up, including the retrospective time period from baseline to 6 months. Comparably, Werner and Ihlebæk [222] showed that 39% of patients with low back pain in 2011 were referred for imaging by GPs in Norway. Likewise, in a recent systematic review of healthcare provided for patients with low back pain, Kamper et al. [91] reported that around 1 in 4 was referred for imaging in family practices. Thus, our finding is more or less in line with previous research. Updated clinical guidelines recommend that imaging should not be routinely used, but rather be reserved for patients for whom the results are likely to alter management [43, 92, 93, 223], also among older people [94]. Evidence suggests that the prevalence of serious pathology as a cause of back pain, for which imaging is indicated, is ≤ 6% in primary care [30, 32, 33, 224]. Moreover, imaging is not recommended as part of the initial assessment for spinal stenosis, but should be reserved for patients being considered for surgery [225-227]. In this context, a rate of 34% seems to indicate an overuse of imaging [223, 228]. This is despite the fact that the prevalence of spinal stenosis and specific pathological causes of back pain increases with age, as illuminated in chapter 2.1.1 [29-33].

Our findings regarding medication use in Paper I differ slightly from previous research. In our study, paracetamol (27-35% of patients) followed by NSAIDs (20-24% of patients) were used most frequently, whereas only a small proportion of patients used opioids (1-2% of patients). Estimates provided by Kamper et al. [91] have suggested that around 20% of patients with low back pain within family practices are recommended paracetamol, 35-40% NSAIDs, and up to 30% opioids. Differences in paracetamol use might be explained by the fact that most studies do not include over-the-counter medication, thus the use of paracetamol is probably underrepresented in the review by Kamper et al. [91]. Differences in NSAIDs use might be explained by the fact that our sample consists of exclusively older people who often have a higher risk of NSAID-related side effects [229, 230]. Differences in opioid use might be explained by the strict opioid prescription regulations in Norway [231]. Updated clinical guidelines recommend pharmacological treatment as an adjunctive option in case of an inadequate response to first-line treatment [43, 92, 97]. Taking possible side effects into account, NSAIDs should be first-line pharmacological treatment. Opioids should be used only in carefully selected patients. Paracetamol is not recommended. In this context, it appears that, unlike paracetamol use, opioid use within Paper I might be in line with clinical guidelines.

Medium- and high-risk patients had a significantly higher degree of healthcare utilization compared to low-risk patients in Paper I. To the best of my knowledge, no previous

study conducted in a usual care setting has described healthcare utilization across patients with different risk profiles, stratified by the SBST. Updated clinical guidelines recommend a stratified healthcare approach [43, 96, 97]. In this context, it is promising to find that low-risk patients have a lower degree of healthcare utilization compared to medium- and high-risk patients.

In Paper II, four modifiable prognostic factors associated with high costs related to healthcare utilization were identified and replicated: pain severity, disability, depression, and physical health-related quality of life. These findings are supported by several previous studies on mainly middle-aged patients with back pain [64, 75, 102, 103, 105, 107, 109] and musculoskeletal disorders [110-112]. Likewise, our results showing no prognostic value of kinesiophobia and our inconsistent results with regards to the prognostic value of radiating pain below the knee are in line with a previous study [75], which showed that these factors were of only minor importance when predicting future costs related to healthcare utilization among mainly middle-aged patients with back pain. Our inconsistent results regarding the prognostic value of mental health-related quality of life and comorbidity are, however, contrary to previous research. In a former study on mainly middle-aged patients with musculoskeletal disorders, mental health-related quality of life was found to be associated with persistent high costs related to healthcare utilization [110]. Furthermore, in a recent systematic review, comorbidity has been pointed out as a consistent prognostic factor of high costs related to healthcare utilization in general [113], and similar conclusions have been drawn in single studies among mainly middle-aged patients with back pain [61] and musculoskeletal disorders [111]. This discrepancy might be explained by the fact that we only included costs related to back pain specific healthcare utilization, whereas other studies have included healthcare costs related to all types of musculoskeletal disorders [61, 111] and healthcare costs in general [113]. To the best of my knowledge, the prognostic value of recovery expectations for high costs related to healthcare utilization has not been reported previously.

Papers III and IV

The iPCQ is a relatively new questionnaire, thus, only 3 methodological studies have been conducted on the iPCQ with which we can compare our results from Papers III and IV [10, 122, 232]. To the best of my knowledge, content validity of the iPCQ has not been tested previously. Our evaluation of content validity showed that all included items were considered to be relevant and to cover the main domains of productivity costs, except compensation

mechanisms and part-time sick leave. Moreover, the comprehensibility of the iPCQ was good. Compensation mechanisms may influence the total value of productivity costs [10, 115, 186]. The extent to which these mechanisms affect final productivity costs remains unclear and there is currently a general agreement that adjusting for them seems premature [10, 115, 186]. However, it would be possible to include items covering part-time sick leave, increasing the usefulness of the iPCQ thereby. In a modified version of the iPCQ, the iPCQ-VR, Beemster et al. [122] showed that part-time sick leave can be reliably measured in Dutch patients with nonspecific musculoskeletal pain.

Our assessment of construct validity confirmed a 3-component solution, which is similar to the original study of the iPCQ [10]. Bouwmans et al. [10] distinguished between the 3 components based on a theoretical rationale, whereas Paper III is the first to confirm this structure based on statistical analysis. The 3 components accounted for as much as 82% of the total variance in the data, the internal consistency was acceptable for all 3 components, and construct validity was supported by hypothesis testing.

In Paper IV, we found that the iPCQ reported occurrence and duration of long-term absenteeism adequately reflected publicly registered absenteeism. However, for adjusted duration (no. of days with complete absenteeism) of long-term absenteeism, the iPCQ overestimated numbers of days with complete sick leave as compared to public registry data. To the best of my knowledge, no previous study has validated iPCQ reported absenteeism against public registry data. Nevertheless, there are studies comparing self-reported and publicly registered absenteeism, albeit with different self-reported questions. Our results regarding occurrence of long-term absenteeism are in line with a previous study on patients with sciatica by Grøvle et al. [233], which also showed an overall agreement of 85% between self-reported and registry data on the occurrence of absenteeism. Likewise, in a cohort study on employees in the Swedish public sector, Voss et al. [234] reported an overall agreement of 74% to 91%. With regards to the duration (no. of days with absenteeism) of long-term absenteeism, our results are in line with a recent meta-analysis, which supports satisfactory agreement between self-reported and registry data on the duration of absenteeism [235]. To the best of my knowledge, no previous study has compared self-reported and registered adjusted duration (no. of days with complete absenteeism) of long-term absenteeism. However, as pointed out in a previous study [122], it seems reasonable that a measuring tool which does not cover part-time sick leave would tend to overestimate the total amount of long-term absenteeism, as well as related costs.

Finally, we found that the reliability of the iPCQ was good, except for item no. 8 (*no. of workdays with disability*). The study by Beemster et al. [122] supports good reliability of items related to long-term absenteeism and low reliability of item no. 8. However, with regards to items related to short-term absenteeism (item no. 4) and presenteeism (item no. 9), our results indicated a higher reliability than was found by Beemster et al. [122]. This discrepancy might be explained by a different time interval between test and retest, as Beemster et al. [122] used an average of 20 days, compared with 3 in our study. Moreover, the study of Kim et al. [232], in which the Korean version of the iPCQ was tested among female outpatients at a gynaecological clinic, supports good reliability of the 3 index scores. To the best of my knowledge, the reliability of items covering productivity costs related to unpaid work (items no. 11 and 12) have not been tested previously.

CONCLUSIONS

Three research gaps were identified in this thesis. Papers I-IV were conducted to bridge these gaps. The following conclusions can be drawn from the included papers:

I. The one-year mean and median total costs of healthcare utilization per patient were estimated at €825 and €364, respectively. The largest cost category was primary care consultations, accounting for 56% of total costs. Imaging rate was 34%. The most frequently used medication was paracetamol (27-35% of patients). Patients in the top 25th percentile accounted for 77% of total costs. Patients with medium and high risk of persistent disabling back pain had a significantly higher degree of healthcare utilization compared to patients with low risk.

II. Four potential modifiable prognostic factors associated with high costs related to healthcare utilization were identified and replicated: pain severity, disability, depression, and physical health-related quality of life.

III. The Norwegian iPCQ showed good measurement properties among patients with musculoskeletal disorders from secondary care in Norway.

IV. The Norwegian iPCQ showed good agreement with public registry data regarding the occurrence and duration of long-term absenteeism among people with musculoskeletal disorders on long-term sick leave in Norway. However, the iPCQ does not cover part-time sick leave and may therefore overestimate the total amount of long-term absenteeism.

IMPLICATIONS

This thesis provides knowledge that can inform our use of scarce healthcare resources and reduce the economic burden of back pain on healthcare systems. Moreover, it contributes knowledge useful for conducting future comprehensive health economic evaluations, in which productivity costs are included. The following implications can be drawn from this thesis:

- In the context of clinical guidelines, decreasing the use of imaging and paracetamol seems to be important for quality improvement in the primary care management of older patients with back pain.
- Pain severity, disability, depression, and physical health-related quality of life are potential target areas for interventions which could reduce high costs related to healthcare utilization among older patients with back pain.
- The iPCQ can be recommended as a useful tool for measuring three important components of productivity costs among people with musculoskeletal disorders: absenteeism, presenteeism, and costs related to unpaid work.

FUTURE PERSPECTIVES

Further studies on healthcare utilization and the related costs among older people with back pain are needed to complement our findings in Paper I. Future studies should preferably be conducted from a societal perspective in order to also gain knowledge on indirect costs related to productivity loss. Although most people above the age of 65-67 years are taking retirement, people are often encouraged to stay in work longer and the pension age might be raised in the future. Therefore, including indirect costs in COI studies on back pain among older people is important. Furthermore, studies with more frequent follow-ups would be beneficial to reduce the risk of recall bias and avoid periods with a lack of data. New, high quality studies would provide valuable knowledge about direct and indirect costs related to back pain among older people. Moreover, potential gaps between clinical guidelines and practice may be found, revealing areas for quality improvement in the management of older people with back pain.

Future research aimed at identifying and replicating modifiable prognostic factors of high costs related to healthcare utilization is warranted to inform and facilitate intervention research that can steer us toward improved use of our scarce healthcare resources and cost reduction related to healthcare utilization. Ideally, future studies should be conducted from a societal perspective including both direct and indirect costs, as indirect costs related to productivity loss are expected to far outweigh direct costs related to healthcare utilization. The economic burden of back pain is substantial. With an aging population, one goal of future back pain research must be to gain knowledge on how to reduce the socio-economic burden of back pain among older people.

Since the iPCQ is a generic instrument, future research should validate the iPCQ among other populations, and criterion validity should be evaluated among people on short-term sick leave (≤ 4 weeks). In addition, it seems appropriate to include and evaluate the validity of items covering part-time sick leave. Future research assessing content validity based upon the updated guidelines from COSMIN [5] is probably needed as well. Continued research on the iPCQ is warranted and expected to improve the toolset needed in health economic evaluations. Valid instruments of productivity loss suited for use in health economic evaluations are needed [20, 37, 118]. In a recent systematic review, the iPCQ was recommended as probably the most suitable instrument for use in health economic evaluations [121].

REFERENCES

1. Dionne, C.E., et al., *A consensus approach toward the standardization of back pain definitions for use in prevalence studies.* Spine (Phila Pa 1976), 2008. **33**(1): p. 95-103.
2. Kamper, S.J., *Bias: Linking Evidence With Practice.* J Orthop Sports Phys Ther, 2018. **48**(8): p. 667-668.
3. Prüss-Üstün, A., et al., *Introduction and methods: assessing the environmental burden of disease at national and local levels (WHO Environmental Burden of Disease Series, No. 1).* 2003, World Health Organization: Geneva.
4. Mokkink, L.B., et al., *The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes.* Journal of clinical epidemiology, 2010. **63**(7): p. 737-745.
5. Terwee, C.B., et al., *COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study.* Qual Life Res, 2018. **27**(5): p. 1159-1170.
6. Kobelt, G., *Health Economics: An Introduction to Economic Evaluation*. Third ed. 2013, United Kingdom: Office of Health Economics.
7. Guinness, L. and V. Wiseman, *Introduction to Health Economics*. second ed. 2011: McGraw-Hill Open University Press.
8. Larg, A. and J.R. Moss, *Cost-of-illness studies: a guide to critical evaluation.* Pharmacoeconomics, 2011. **29**(8): p. 653-71.
9. Jo, C., *Cost-of-illness studies: concepts, scopes, and methods.* Clin Mol Hepatol, 2014. **20**(4): p. 327-37.
10. Bouwmans, C., et al., *The iMTA Productivity Cost Questionnaire: A Standardized Instrument for Measuring and Valuing Health-Related Productivity Losses.* Value Health, 2015. **18**(6): p. 753-8.
11. Field-Fote, E.E., *Mediators and Moderators, Confounders and Covariates: Exploring the Variables That Illuminate or Obscure the "Active Ingredients" in Neurorehabilitation.* J Neurol Phys Ther, 2019. **43**(2): p. 83-84.
12. Riley, R.D., et al., *Prognosis Research Strategy (PROGRESS) 2: prognostic factor research.* PLoS Med, 2013. **10**(2): p. e1001380.
13. Hemingway, H., et al., *Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes.* BMJ, 2013. **346**: p. e5595.
14. Steyerberg, E.W., et al., *Prognosis Research Strategy (PROGRESS) 3: prognostic model research.* PLoS Med, 2013. **10**(2): p. e1001381.
15. Hingorani, A.D., et al., *Prognosis research strategy (PROGRESS) 4: stratified medicine research.* BMJ, 2013. **346**: p. e5793.
16. Vos, T., et al., *Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016.* Lancet, 2017. **390**(10100): p. 1211-1259.
17. Chen, S., et al., *Global, regional and national burden of low back pain 1990-2019: A systematic analysis of the Global Burden of Disease study 2019.* J Orthop Translat, 2022. **32**: p. 49-58.
18. van Tulder, M., B. Koes, and C. Bombardier, *Low back pain.* Best Pract Res Clin Rheumatol, 2002. **16**(5): p. 761-75.
19. Dagenais, S., J. Caro, and S. Haldeman, *A systematic review of low back pain cost of illness studies in the United States and internationally.* Spine J, 2008. **8**(1): p. 8-20.

20. Zemedikun, D.T., et al., *Methodological considerations in the assessment of direct and indirect costs of back pain: A systematic scoping review.* PLoS One, 2021. **16**(5): p. e0251406.
21. Paeck, T., et al., *Are older adults missing from low back pain clinical trials? A systematic review and meta-analysis.* Arthritis Care Res (Hoboken), 2014. **66**(8): p. 1220-6.
22. Bressler, H.B., et al., *The Prevalene of Low Back Pain in the Elderly: A Systematic Review of the Literature.* Spine, 1999. **24**(17): p. 1813-1819.
23. Scheele, J., et al., *Course of back complaints in older adults: a systematic literature review.* European Journal of PHysical Rehabilitation Medicine, 2012. **48**(3): p. 379-386.
24. van der Windt, D.A. and K.M. Dunn, *Low back pain research--future directions.* Best Pract Res Clin Rheumatol, 2013. **27**(5): p. 699-708.
25. Maher, C., M. Underwood, and R. Buchbinder, *Non-specific low back pain.* The Lancet, 2017. **389**(10070): p. 736-747.
26. Westert, G.P., et al., *Patterns of comorbidity and the use of health services in the Dutch population.* Eur J Public Health, 2001. **11**(4): p. 365-72.
27. Thomas, E., et al., *The effect of age on the onset of pain interference in a general population of older adults: prospective findings from the North Staffordshire Osteoarthritis Project (NorStOP).* Pain, 2007. **129**(1-2): p. 21-7.
28. Docking, R.E., et al., *Epidemiology of back pain in older adults: prevalence and risk factors for back pain onset.* Rheumatology (Oxford), 2011. **50**(9): p. 1645-53.
29. Wong, A.Y., J. Karppinen, and D. Samartzis, *Low back pain in older adults: risk factors, management options and future directions.* Scoliosis Spinal Disord, 2017. **12**: p. 14.
30. Henschke, N., et al., *Prevalence of and screening for serious spinal pathology in patients presenting to primary care settings with acute low back pain.* Arthritis Rheum, 2009. **60**(10): p. 3072-80.
31. Street, K.J., S.G. White, and A.C. Vandal, *Clinical prevalence and population incidence of serious pathologies among patients undergoing magnetic resonance imaging for low back pain.* Spine J, 2020. **20**(1): p. 101-111.
32. Hartvigsen, J., et al., *What low back pain is and why we need to pay attention.* Lancet, 2018. **391**(10137): p. 2356-2367.
33. Enthoven, W.T., et al., *Prevalence and "Red Flags" Regarding Specified Causes of Back Pain in Older Adults Presenting in General Practice.* Physical Therapy, 2016. **96**(3): p. 305-312.
34. Scheele, J., et al., *Back complaints in the elders (BACE); design of cohort studies in primary care: an international consortium.* BMC Musculoskelet Disord, 2011. **12**: p. 193.
35. Grotle, M., *BACk Pain in Elders in Norway (BACE-N).* 2020, ClinicalTrials.gov: https://clinicaltrials.gov/ct2/show/NCT04261309.
36. Krol, M., et al., *A noticeable difference? Productivity costs related to paid and unpaid work in economic evaluations on expensive drugs.* Eur J Health Econ, 2016. **17**(4): p. 391-402.
37. Tang, K., *Estimating productivity costs in health economic evaluations: a review of instruments and psychometric evidence.* Pharmacoeconomics, 2015. **33**(1): p. 31-48.
38. Buchbinder, R., et al., *The Lancet Series call to action to reduce low value care for low back pain: an update.* Pain, 2020. **161 Suppl 1**: p. S57-S64.
39. Hoy, D., et al., *A systematic review of the global prevalence of low back pain.* Arthritis & Rheumatism, 2012. **64**(6): p. 2028-2037.

40. Dionne, C.E., K.M. Dunn, and P.R. Croft, *Does back pain prevalence really decrease with increasing age? A systematic review.* Age and ageing, 2006. **35**(3): p. 229-234.
41. Buchbinder, R., et al., *Placing the global burden of low back pain in context.* Best Pract Res Clin Rheumatol, 2013. **27**(5): p. 575-89.
42. Briggs, A.M., et al., *Musculoskeletal Health Conditions Represent a Global Threat to Healthy Aging: A Report for the 2015 World Health Organization World Report on Ageing and Health.* Gerontologist, 2016. **56 Suppl 2**: p. S243-55.
43. Foster, N.E., et al., *Prevention and treatment of low back pain: evidence, challenges, and promising directions.* Lancet, 2018. **391**(10137): p. 2368-2383.
44. Hartvigsen, L., et al., *Leg pain location and neurological signs relate to outcomes in primary care patients with low back pain.* BMC Musculoskelet Disord, 2017. **18**(1): p. 133.
45. Tomkins-Lane, C., et al., *ISSLS Prize Winner: Consensus on the Clinical Diagnosis of Lumbar Spinal Stenosis: Results of an International Delphi Study.* Spine (Phila Pa 1976), 2016. **41**(15): p. 1239-1246.
46. Kalichman, L., et al., *Spinal stenosis prevalence and association with symptoms: the Framingham Study.* The Spine Journal, 2009. **9**(7): p. 545-550.
47. Brinjikji, W., et al., *Systematic literature review of imaging features of spinal degeneration in asymptomatic populations.* AJNR Am J Neuroradiol, 2015. **36**(4): p. 811-6.
48. Boden, S.D., et al., *Abnormal magnetic-resonance scans of the lumbar spine in asymptomatic subjects. A prospective investigation.* J Bone Joint Surg Am, 1990. **72**(3): p. 403-8.
49. Waterloo, S., et al., *Prevalence of vertebral fractures in women and men in the population-based Tromso Study.* BMC Musculoskelet Disord, 2012. **13**: p. 3.
50. Waddell, G., *1987 Volvo award in clinical sciences. A new clinical model for the treatment of low-back pain.* Spine (Phila Pa 1976), 1987. **12**(7): p. 632-44.
51. Maher, C. and G. Ferreira, *Time to reconsider what Global Burden of Disease studies really tell us about low back pain.* Ann Rheum Dis, 2022. **81**(3): p. 306-308.
52. Vos, T., et al., *Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019.* Lancet, 2020. **396**(10258): p. 1204-1222.
53. World Health Orginazation, *Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks*. 2009, WHO Press: Geneva.
54. Deyo, R.A., S.K. Mirza, and B.I. Martin, *Back pain prevalence and visit rates: estimates from U.S. national surveys, 2002.* Spine (Phila Pa 1976), 2006. **31**(23): p. 2724-7.
55. Licciardone, J.C., *The epidemiology and medical management of low back pain during ambulatory medical care visits in the United States.* Osteopath Med Prim Care, 2008. **2**: p. 11.
56. Kinge, J.M., et al., *Musculoskeletal disorders in Norway: prevalence of chronicity and use of primary and specialist health care services.* BMC Musculoskeletal Disorders, 2015. **16**(1): p. 1-9.
57. Werner, E.L.I.A., *Kunnskap, praksis og holdninger til rygglidelser hos leger, fysioterapeuter og kiropraktorer.* Tidsskr Nor Laegeforen 2005. **125**(13): p. 1794-7.
58. Vos, T., et al., *Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015.* Lancet, 2016. **388**(10053): p. 1545-1602.

59. Macfarlane, G.J., et al., *The prevalence and management of low back pain across adulthood: results from a population-based cross-sectional study (the MUSICIAN study).* Pain, 2012. **153**(1): p. 27-32.
60. Joud, A., I.F. Petersson, and M. Englund, *Low back pain: epidemiology of consultations.* Arthritis Care Res (Hoboken), 2012. **64**(7): p. 1084-8.
61. Ritzwoller, D.P., et al., *The association of comorbidities, utilization and costs for patients identified with low back pain.* BMC Musculoskelet Disord, 2006. **7**: p. 72.
62. Bevan, S., et al., *Fit For work? Musculoskeletal disorders in the European workforce*. 2009, The Work Foundation: fit For work Europe.
63. Maniadakis, N. and A. Gray, *The economic burden of back pain in the UK.* Pain, 2000. **84**(1): p. 95-103.
64. Wenig, C.M., et al., *Costs of back pain in Germany.* Eur J Pain, 2009. **13**(3): p. 280-6.
65. Ekman, M., O. Johnell, and L. Lidgren, *The economic cost of low back pain in Sweden in 2001.* Acta Orthop, 2005. **76**(2): p. 275-84.
66. Rice, D.P., *Cost of illness studies: what is good about them?* Inj Prev, 2000. **6**(3): p. 177-9.
67. Buchbinder, R., et al., *Low back pain: a call for action.* Lancet, 2018. **391**(10137): p. 2384-2388.
68. Montgomery, W., et al., *The economic and humanistic costs of chronic lower back pain in Japan.* Clinicoecon Outcomes Res, 2017. **9**: p. 361-371.
69. Depont, F., et al., *Medical and non-medical direct costs of chronic low back pain in patients consulting primary care physicians in France.* Fundam Clin Pharmacol, 2010. **24**(1): p. 101-8.
70. Hong, J., et al., *Costs associated with treatment of chronic low back pain: an analysis of the UK General Practice Research Database.* Spine (Phila Pa 1976), 2013. **38**(1): p. 75-82.
71. Walker, B.F., R. Muller, and W.D. Grant, *Low back pain in Australian adults: the economic burden.* Asia Pac J Public Health, 2003. **15**(2): p. 79-87.
72. van Zundert, J. and M. van Kleef, *Low back pain: from algorithm to cost-effectiveness?* Pain Pract, 2005. **5**(3): p. 179-89.
73. Coyte, P.C., et al., *The economic cost of musculoskeletal disorders in Canada.* Arthritis Care Res, 1998. **11**(5): p. 315-25.
74. Hemmila, H.M., *Quality of life and cost of care of back pain patients in Finnish general practice.* Spine (Phila Pa 1976), 2002. **27**(6): p. 647-53.
75. Becker, A., et al., *Low back pain in primary care: costs of care and prediction of future health care utilization.* Spine (Phila Pa 1976), 2010. **35**(18): p. 1714-20.
76. Itoh, H., F. Kitamura, and K. Yokoyama, *Estimates of annual medical costs of work-related low back pain in Japan.* Ind Health, 2013. **51**(5): p. 524-9.
77. Boonen, A., et al., *Large differences in cost of illness and wellbeing between patients with fibromyalgia, chronic low back pain, or ankylosing spondylitis.* Ann Rheum Dis, 2005. **64**(3): p. 396-402.
78. Lambeek, L.C., et al., *The trend in total cost of back pain in The Netherlands in the period 2002 to 2007.* Spine (Phila Pa 1976), 2011. **36**(13): p. 1050-8.
79. van Tulder, M.W., B.W. Koes, and L.M. Bouter, *A cost-of-illness study of back pain in The Netherlands.* Pain, 1995. **62**(2): p. 233-240.
80. Alonso-Garcia, M. and A. Sarria-Santamera, *The Economic and Social Burden of Low Back Pain in Spain: A National Assessment of the Economic and Social Impact of Low Back Pain in Spain.* Spine (Phila Pa 1976), 2020. **45**(16): p. E1026-E1032.

81. Ekman, M., et al., *Burden of illness of chronic low back pain in Sweden: a cross-sectional, retrospective study in primary care setting.* Spine (Phila Pa 1976), 2005. **30**(15): p. 1777-85.
82. Hansson, E.K. and T.H. Hansson, *The costs for persons sick-listed more than one month because of low back or neck problems. A two-year prospective study of Swedish patients.* Eur Spine J, 2005. **14**(4): p. 337-45.
83. Jonsson, D. and M. Husberg, *Socioeconomic costs of rheumatic diseases. Implications for technology assessment.* Int J Technol Assess Health Care, 2000. **16**(4): p. 1193-200.
84. Olafsson, G., et al., *Cost of low back pain: results from a national register study in Sweden.* Eur Spine J, 2018. **27**(11): p. 2875-2881.
85. Wieser, S., et al., *Cost of low back pain in Switzerland in 2005.* Eur J Health Econ, 2011. **12**(5): p. 455-67.
86. Icatasiotlu, A., et al., *Burden of Chronic Low Back Pain in the Turkish Population.* Turk J Phys Med Rehab, 2015. **61**: p. 58-64.
87. Yumusakhuylu, Y., et al., *The Economic Cost of Chronic Low Back Pain.* Bozok Med J, 2018. **8**(1): p. 66-74.
88. Luo, X., et al., *Estimates and patterns of direct health care expenditures among individuals with back pain in the United States.* Spine (Phila Pa 1976), 2004. **29**(1): p. 79-86.
89. Martin, B.I., et al., *Expenditures and health status among adults with back and neck problems.* JAMA, 2008. **299**(6): p. 656-64.
90. Smith, M., et al., *Aging baby boomers and the rising cost of chronic back pain: secular trend analysis of longitudinal Medical Expenditures Panel Survey data for years 2000 to 2007.* J Manipulative Physiol Ther, 2013. **36**(1): p. 2-11.
91. Kamper, S.J., et al., *What is usual care for low back pain? A systematic review of health care provided to patients with low back pain in family practice and emergency departments.* Pain, 2020. **161**(4): p. 694-702.
92. Corp, N., et al., *Evidence-based treatment recommendations for neck and low back pain across Europe: A systematic review of guidelines.* Eur J Pain, 2021. **25**(2): p. 275-295.
93. O'Connell, N.E., et al., *Clinical guidelines for low back pain: A critical review of consensus and inconsistencies across three major guidelines.* Best Practice & Research Clinical Rheumatology, 2016. **30**(6): p. 968-980.
94. Jarvik, J.G., et al., *Association of early imaging for back pain with clinical outcomes in older adults.* JAMA, 2015. **313**(11): p. 1143-53.
95. Hill, J.C., et al., *A primary care back pain screening tool: identifying patient subgroups for initial treatment.* Arthritis Rheum, 2008. **59**(5): p. 632-41.
96. Traeger, A.C., et al., *Care for low back pain: can health systems deliver?* Bull World Health Organ, 2019. **97**(6): p. 423-433.
97. NICE, *Low back pain and sciatica in over 16s: assessment and management*. 2016, National Institute for Clinical Excellence: London.
98. Hall, J.A., et al., *The STarT Back stratified care model for nonspecific low back pain: a model-based evaluation of long-term cost-effectiveness.* Pain, 2021. **162**(3): p. 702-710.
99. Lin, I., et al., *What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review.* Br J Sports Med, 2020. **54**(2): p. 79-86.
100. Vlaeyen, J.W.S., et al., *Low back pain.* Nat Rev Dis Primers, 2018. **4**(1): p. 52.

101. Andersen, R.M., *National health surveys and the behavioral model of health services use.* Med Care, 2008. **46**(7): p. 647-53.
102. Mutubuki, E.N., et al., *Predictive factors of high societal costs among chronic low back pain patients.* Eur J Pain, 2019.
103. Engel, C.C., M. von Korff, and W.J. Katon, *Back pain in primary care: predictors of high health-care costs.* Pain, 1996. **65**(2-3): p. 197-204.
104. Mannion, A.F., S. Wieser, and A. Elfering, *Association between beliefs and care-seeking behavior for low back pain.* Spine (Phila Pa 1976), 2013. **38**(12): p. 1016-25.
105. Lim, K.L., P. Jacobs, and S. Klarenbach, *A population-based analysis of healthcare utilization of persons with back disorders: results from the Canadian Community Health Survey 2000-2001.* Spine (Phila Pa 1976), 2006. **31**(2): p. 212-8.
106. Hirsch, O., et al., *Low back pain patient subgroups in primary care: pain characteristics, psychosocial determinants, and health care utilization.* Clin J Pain, 2014. **30**(12): p. 1023-32.
107. Keeley, P., et al., *Psychosocial predictors of health-related quality of life and health service utilisation in people with chronic low back pain.* Pain, 2008. **135**(1-2): p. 142-50.
108. Stewart, W.F., et al., *Patterns of health care utilization for low back pain.* J Pain Res, 2015. **8**: p. 523-35.
109. Ferreira, M., et al., *Factors defining care-seeking in low back pain - A meta-analysis of population based surveys.* European Journal of Pain, 2010. **14**(7): p. 747.e1-747.e7.
110. Lentz, T.A., et al., *Factors associated with persistently high-cost health care utilization for musculoskeletal pain.* PLoS One, 2019. **14**(11): p. e0225125.
111. Lentz, T.A., J.M. Beneciuk, and S.Z. George, *Prediction of healthcare utilization following an episode of physical therapy for musculoskeletal pain.* Bmc Health Services Research, 2018. **18**.
112. Budtz, C.R., S. Mose, and D.H. Christiansen, *Socio-demographic, clinical and psychological predictors of healthcare utilization among patients with musculoskeletal disorders: a prospective cohort study.* BMC Health Serv Res, 2020. **20**(1): p. 239.
113. Wammes, J.J.G., et al., *Systematic review of high-cost patients' characteristics and healthcare utilisation.* Bmj Open, 2018. **8**(9).
114. Sharma, D., et al., *National Healthcare Economic Evaluation Guidelines: A Cross-Country Comparison.* Pharmacoecon Open, 2021. **5**(3): p. 349-364.
115. Krol, M., W. Brouwer, and F. Rutten, *Productivity costs in economic evaluations: past, present, future.* Pharmacoeconomics, 2013. **31**(7): p. 537-49.
116. Krol, M. and W. Brouwer, *How to estimate productivity costs in economic evaluations.* Pharmacoeconomics, 2014. **32**(4): p. 335-44.
117. Zhang, W., N. Bansback, and A.H. Anis, *Measuring and valuing productivity loss due to poor health: A critical review.* Soc Sci Med, 2011. **72**(2): p. 185-92.
118. Mattke, S., et al., *A review of methods to measure health-related productivity loss.* Am J Manag Care, 2007. **13**(4): p. 211-7.
119. Lofland, J.H., L. Pizzi, and K.D. Frick, *A review of health-related workplace productivity loss instruments.* Pharmacoeconomics, 2004. **22**(3): p. 165-84.
120. Prasad, M., et al., *A review of self-report instruments measuring health-related work productivity: a patient-reported outcomes perspective.* Pharmacoeconomics, 2004. **22**(4): p. 225-44.
121. Hubens, K., et al., *Measurement Instruments of Productivity Loss of Paid and Unpaid Work: A Systematic Review and Assessment of Suitability for Health Economic Evaluations From a Societal Perspective.* Value Health, 2021. **24**(11): p. 1686-1699.

122. Beemster, T.T., et al., *Test-Retest Reliability, Agreement and Responsiveness of Productivity Loss (iPCQ-VR) and Healthcare Utilization (TiCP-VR) Questionnaires for Sick Workers with Chronic Musculoskeletal Pain.* J Occup Rehabil, 2019. **29**(1): p. 91-103.
123. Tveter, A.T., et al., *Risk assessment for prolonged sickness absence due to musculoskeletal disorders: protocol for a prospective cohort study.* BMC Musculoskelet Disord, 2020. **21**(1): p. 326.
124. Mokkink, L.B., et al., *The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study.* Qual Life Res, 2010. **19**(4): p. 539-49.
125. Mokkink, L.B., et al., *COSMIN Study Design checklist for Patient-reported outcome measurement instruments*. 2019: www.cosmin.nl.
126. Killingmo, R.M., et al., *Statistical analysis plan (SAP) for: Healthcare utilization and related costs among older people seeking primary care due to back pain: a prospective cohort study with one year of follow up*. 2021, ClinicalTrials.gov: https://clinicaltrials.gov/ProvidedDocs/09/NCT04261309/SAP_001.pdf.
127. Killingmo, R.M., et al., *Statistical analysis plan (SAP) for: Modifiable prognostic factors of high cost related to healthcare utilization among older people seeking primary care with a new episode of back pain - an identification and replication study*. 2021, ClinicalTrials.gov: https://clinicaltrials.gov/ProvidedDocs/09/NCT04261309/SAP_002.pdf.
128. Beaton, D.E., et al., *Guidelines for the process of cross-cultural adaptation of self-report measures.* Spine (Phila Pa 1976), 2000. **25**(24): p. 3186-91.
129. Guillemin, F., C. Bombardier, and D. Beaton, *Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines.* J Clin Epidemiol, 1993. **46**(12): p. 1417-32.
130. Bouwmans, C., et al., *Manual: iMTA Productivity Costs Questionnaire*. 2013, Institute for Medical Technology Assessment: www.imta.nl/questionnaires.
131. Kanters, T. *Questionnaires for the measurement of costs in economic evaluations*. [cited 2018.
132. Bouwmans, C.H.-v.R., L.; Koopmanschap, M.; Krol, M.; Severens, H.; Brouwer, W., *Manual - iMTA Productivity Costs Questionnaire*, I.f.M.T.A.E.U. Rotterdam, Editor. 2013: www.imta.nl/questionnaires.
133. Strand, L.I., et al., *The Short-Form McGill Pain Questionnaire as an outcome measure: test-retest reliability and responsiveness to change.* Eur J Pain, 2008. **12**(7): p. 917-25.
134. Melzack, R., *The McGill Pain Questionnaire: major properties and scoring methods.* Pain, 1975. **1**(3): p. 277-299.
135. Von Korff, M., M.P. Jensen, and P. Karoly, *Assessing global pain severity by self-report in clinical and health services research.* Spine (Phila Pa 1976), 2000. **25**(24): p. 3140-51.
136. Ferreira-Valente, M.A., J.L. Pais-Ribeiro, and M.P. Jensen, *Validity of four pain intensity rating scales.* Pain, 2011. **152**(10): p. 2399-2404.
137. Strong, J., R. Ashton, and D. Chant, *Pain intensity measurement in chronic low back pain.* Clin J Pain, 1991. **7**(3): p. 209-18.
138. Grotle, M., J.I. Brox, and N.K. Vollestad, *Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain.* Spine (Phila Pa 1976), 2004. **29**(21): p. E492-501.

139. Roland, M. and R. Morris, *A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain.* Spine (Phila Pa 1976), 1983. **8**(2): p. 141-4.
140. Grotle, M., J.I. Brox, and N.K. Vollestad, *Functional status and disability questionnaires: what do they assess? A systematic review of back-specific outcome questionnaires.* Spine (Phila Pa 1976), 2005. **30**(1): p. 130-40.
141. Grotle, M., J.I. Brox, and N.K. Vollestad, *Cross-cultural adaptation of the Norwegian versions of the Roland-Morris Disability Questionnaire and the Oswestry Disability Index.* J Rehabil Med, 2003. **35**(5): p. 241-7.
142. Sangha, O., et al., *The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research.* Arthritis Rheum, 2003. **49**(2): p. 156-63.
143. Stolwijk, C., et al., *Aspects of validity of the self-administered comorbidity questionnaire in patients with ankylosing spondylitis.* Rheumatology (Oxford), 2014. **53**(6): p. 1054-64.
144. Ware, J.E., Jr. and C.D. Sherbourne, *The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection.* Med Care, 1992. **30**(6): p. 473-83.
145. Bergman, S., et al., *Health status as measured by SF-36 reflects changes and predicts outcome in chronic musculoskeletal pain: a 3-year follow up study in the general population.* Pain, 2004. **108**(1-2): p. 115-23.
146. Beaton, D.E., S. Hogg-Johnson, and C. Bombardier, *Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders.* J Clin Epidemiol, 1997. **50**(1): p. 79-93.
147. Radloff, L.S., *The CES-D scale: a Self-Report Depression Scale for Research in the General Population.* Applied psychological measurement, 1977. (1(3)): p. 385-355.
148. Waddell, G., et al., *A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability.* Pain, 1993. **52**(2): p. 157-68.
149. Crombez, G., et al., *Pain-related fear is more disabling than pain itself: evidence on the role of pain-related fear in chronic back pain disability.* Pain, 1999. **80**(1-2): p. 329-39.
150. Pfingsten, M., et al., *Validation of the German version of the Fear-Avoidance Beliefs Questionnaire (FABQ).* Eur J Pain, 2000. **4**(3): p. 259-66.
151. Grotle, M., J.I. Brox, and N.K. Vollestad, *Reliability, validity and responsiveness of the fear-avoidance beliefs questionnaire: methodological aspects of the Norwegian version.* J Rehabil Med, 2006. **38**(6): p. 346-53.
152. Vigdal Ø. N.; Storheim K., G.M., *With what accuracy can we predict persistent disability in sciatica patients with self-reported screening tools?* 2014, Oslo University College.
153. Bier, J.D., et al., *Validity and Reproducibility of the STarT Back Tool (Dutch Version) in Patients With Low Back Pain in Primary Care Settings.* Phys Ther, 2017. **97**(5): p. 561-570.
154. Mehling, W.E., et al., *Can a back pain screening tool help classify patients with acute pain into risk levels for chronic pain?* Eur J Pain, 2015. **19**(3): p. 439-46.
155. Suri, P., et al., *Predictive Validity of the STarT Back Tool for Risk of Persistent Disabling Back Pain in a U.S. Primary Care Setting.* Arch Phys Med Rehabil, 2018. **99**(8): p. 1533-1539 e2.
156. Verhagen, A.P., et al., *Red flags presented in current low back pain guidelines: a review.* Eur Spine J, 2016. **25**(9): p. 2788-802.

157. Kjønø, L.G., et al., *Assessing physical workload among people with musculoskeletal disorders: validity and reliability of the physical workload questionnaire.* BMC Musculoskelet Disord, 2022. **23**(282).
158. Bot, S.D., et al., *Internal consistency and validity of a new physical workload questionnaire.* Occup Environ Med, 2004. **61**(12): p. 980-6.
159. Dallner, M., et al., *Validation of the general Nordicquestionnaire (QPSNordic) for psychological and social factors at work* Copenhagen: Nordic Council of Ministers, 2000. **Nord 2000:12**.
160. Wannstrom, I., et al., *Psychometric properties of scales in the General Nordic Questionnaire for Psychological and Social Factors at Work (QPS): confirmatory factor analysis and prediction of certified long-term sickness absence.* Scand J Psychol, 2009. **50**(3): p. 231-44.
161. Rattay, P., et al., *[Utilization of outpatient and inpatient health services in Germany: results of the German Health Interview and Examination Survey for Adults (DEGS1)].* Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz, 2013. **56**(5-6): p. 832-44.
162. Chechulin, Y., et al., *Predicting patients with high risk of becoming high-cost healthcare users in Ontario (Canada).* Healthc Policy, 2014. **9**(3): p. 68-79.
163. Hoebel, J., et al., *Socioeconomic Status and Use of Outpatient Medical Care: The Case of Germany.* PLoS One, 2016. **11**(5): p. e0155982.
164. Fritz, J.M., J. Kim, and J. Dorius, *Importance of the type of provider seen to begin health care for a new episode low back pain: associations with future utilization and costs.* J Eval Clin Pract, 2016. **22**(2): p. 247-52.
165. Kigozi, J., et al., *Construct validity and responsiveness of the single-item presenteeism question in patients with lower back pain for the measurement of presenteeism.* Spine (Phila Pa 1976), 2014. **39**(5): p. 409-16.
166. Lamers, L.M., et al., *The relationship between productivity and health-related quality of life: an empirical exploration in persons with low back pain.* Qual Life Res, 2005. **14**(3): p. 805-13.
167. Canjuga, M., et al., *Correlates of short- and long-term absence due to musculoskeletal disorders.* Occup Med (Lond), 2010. **60**(5): p. 358-61.
168. Steenstra, I.A., et al., *Prognostic factors for duration of sick leave in patients sick listed with acute low back pain: a systematic review of the literature.* Occup Environ Med, 2005. **62**(12): p. 851-60.
169. Zhang, W., et al., *Measuring time input loss among patients with rheumatoid arthritis: validity and reliability of the Valuation of Lost Productivity questionnaire.* J Occup Environ Med, 2011. **53**(5): p. 530-6.
170. Lardon, A., et al., *Predictors of disability and absenteeism in workers with non-specific low back pain: A longitudinal 15-month study.* Appl Ergon, 2018. **68**: p. 176-185.
171. Koopmanschap, M.A., *PRODISQ: a modular questionnaire on productivity and disease for economic evaluation studies.* Expert Rev Pharmacoecon Outcomes Res, 2005. **5**(1): p. 23-8.
172. Ijzelenberg, W., D. Molenaar, and A. Burdorf, *Different risk factors for musculoskeletal complaints and musculoskeletal sickness absence.* Scand J Work Environ Health, 2004. **30**(1): p. 56-63.
173. d'Errico, A., et al., *Low back pain and associated presenteeism among hospital nursing staff.* J Occup Health, 2013. **55**(4): p. 276-83.

174. Kresal, F., et al., *Psychosocial Risk Factors for Low Back Pain and Absenteeism among Slovenian Professional Drivers.* Cent Eur J Public Health, 2017. **25**(2): p. 135-140.
175. Janssens, H., et al., *Association between psychosocial characteristics of work and presenteeism: A cross-sectional study.* Int J Occup Med Environ Health, 2016. **29**(2): p. 331-44.
176. von Elm, E., et al., *The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies.* PLoS Med, 2007. **4**(10): p. e296.
177. McShane, L.M., et al., *REporting recommendations for tumour MARKer prognostic studies (REMARK).* Br J Cancer, 2005. **93**(4): p. 387-91.
178. Terwee, C.B., et al., *Quality criteria were proposed for measurement properties of health status questionnaires.* J Clin Epidemiol, 2007. **60**(1): p. 34-42.
179. van Ginkel, J.R., et al., *Rebutting Existing Misconceptions About Multiple Imputation as a Method for Handling Missing Data.* J Pers Assess, 2020. **102**(3): p. 297-308.
180. Kanters, T.A., et al., *Update of the Dutch manual for costing studies in health care.* PLoS One, 2017. **12**(11): p. e0187477.
181. Box, G.E. and P.W. Tidwell, *Transformation of the independent variables.* Technometrics, 1962. **4**: p. 531-550.
182. Kent, P., et al., *A conceptual framework for prognostic research.* BMC Med Res Methodol, 2020. **20**(1): p. 172.
183. Pallant, J., *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. 6th ed. 2016, Sydney, Australia: Allen & Unwin.
184. Nunnally, J.C. and I.H. Bernstein, *Psychometric theory*. 3 ed. 1994, New York: McGraw-Hill.
185. Andresen, E.M., *Criteria for assessing the tools of disability outcomes research.* Arch Phys Med Rehabil, 2000. **81**(12 Suppl 2): p. S15-20.
186. Zhang, W. and A.H. Anis, *Health-related productivity loss: NICE to recognize soon, good to discuss now.* Pharmacoeconomics, 2014. **32**(5): p. 425-7.
187. Altman, D.G., *Practical Statistics for Medical Research* 1991, London, UK: Chapman and Hall.
188. Johnston, K.M., et al., *Methods of sample size calculation in descriptive retrospective burden of illness studies.* BMC Med Res Methodol, 2019. **19**(1): p. 9.
189. Harrell, F.E., Jr., et al., *Regression modelling strategies for improved prognostic prediction.* Stat Med, 1984. **3**(2): p. 143-52.
190. Harrell, F.E., Jr., K.L. Lee, and D.B. Mark, *Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors.* Stat Med, 1996. **15**(4): p. 361-87.
191. Steyerberg, E.W., et al., *Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets.* Med Decis Making, 2001. **21**(1): p. 45-56.
192. Steyerberg, E.W., et al., *Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets.* Stat Med, 2000. **19**(8): p. 1059-79.
193. Ambler, G., A.R. Brady, and P. Royston, *Simplifying a prognostic model: a simulation study based on clinical data.* Stat Med, 2002. **21**(24): p. 3803-22.
194. Moons, K.G., et al., *Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist.* PLoS Med, 2014. **11**(10): p. e1001744.

195. Pavlou, M., et al., *Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events.* Stat Med, 2016. **35**(7): p. 1159-77.
196. Cowley, L.E., et al., *Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature.* Diagn Progn Res, 2019. **3**: p. 16.
197. Pavlou, M., et al., *How to develop a more accurate risk prediction model when there are few events.* BMJ, 2015. **351**: p. h3868.
198. Hayden, J.A., et al., *Assessing bias in studies of prognostic factors.* Ann Intern Med, 2013. **158**(4): p. 280-6.
199. Husereau, D., et al., *Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement.* Int J Technol Assess Health Care, 2013. **29**(2): p. 117-22.
200. Mokkink, L.B., et al., *COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures.* Qual Life Res, 2018. **27**(5): p. 1171-1179.
201. Johnson, T.P. and J.S. Wislar, *Response rates and nonresponse errors in surveys.* JAMA, 2012. **307**(17): p. 1805-6.
202. Franklin, M., et al., *An Educational Review About Using Cost Data for the Purpose of Cost-Effectiveness Analysis.* Pharmacoeconomics, 2019. **37**(5): p. 631-643.
203. Leurent, B., M. Gomes, and J.R. Carpenter, *Missing data in trial-based cost-effectiveness analysis: An incomplete journey.* Health Econ, 2018. **27**(6): p. 1024-1040.
204. Briggs, A., et al., *Missing... presumed at random: cost-analysis of incomplete data.* Health Econ, 2003. **12**(5): p. 377-92.
205. Petrou, S., et al., *The accuracy of self-reported healthcare resource utilization in health economic studies.* Int J Technol Assess Health Care, 2002. **18**(3): p. 705-10.
206. Icks, A., et al., *Agreement found between self-reported and health insurance data on physician visits comparing different recall lengths.* J Clin Epidemiol, 2017. **82**: p. 167-172.
207. Short, M.E., et al., *How accurate are self-reports? Analysis of self-reported health care utilization and absence when compared with administrative data.* J Occup Environ Med, 2009. **51**(7): p. 786-96.
208. Hunger, M., et al., *Official statistics and claims data records indicate non-response and recall bias within survey-based estimates of health care utilization in the older population.* BMC Health Serv Res, 2013. **13**: p. 1.
209. Loge, J.H. and S. Kaasa, *Short form 36 (SF-36) health survey: normative data from the general Norwegian population.* Scand J Soc Med, 1998. **26**(4): p. 250-8.
210. Vigdal, Ø.N., et al., *One-year clinical course of back-related disability, and prognostic value of comorbidity on disability in older adults visiting primary care with a new episode of back pain*. 2022 submitted manuscript.
211. Beneciuk, J.M., J.M. Fritz, and S.Z. George, *The STarT Back Screening Tool for prediction of 6-month clinical outcomes: relevance of change patterns in outpatient physical therapy settings.* J Orthop Sports Phys Ther, 2014. **44**(9): p. 656-64.
212. Rothman, K.J., *Six persistent research misconceptions.* J Gen Intern Med, 2014. **29**(7): p. 1060-4.
213. Slagsvold, B.V., M.; Daatland, S. O.; Hagestad, G.; Hansen, T.; Herlofson K.; Koløen, K; Solem, P. E. , *Life-course, ageing and generations in Norway:the NorLAG study.* Norsk Epidemiologi, 2012. **22**(2).
214. Vigdal, O.N., et al., *Characteristics of older adults with back pain associated with choice of first primary care provider: a cross-sectional analysis from the BACE-N cohort study.* BMJ Open, 2021. **11**(9): p. e053229.

215. Grotle, M., et al., *Prognostic factors in first-time care seekers due to acute low back pain.* European Journal of Pain, 2007. **11**(3): p. 290-298.
216. Nordstoga, A.L., et al., *Improvement in Work Ability, Psychological Distress and Pain Sites in Relation to Low Back Pain Prognosis A Longitudinal Observational Study in Primary Care.* Spine, 2019. **44**(7): p. E423-E429.
217. Nations, U., *World Population Ageing 2019* 2020, Department of Economic and Social Affairs, Population Division: New York.
218. Hofman, A., et al., *Determinants of disease and disability in the elderly: the Rotterdam Elderly Study.* Eur J Epidemiol, 1991. **7**(4): p. 403-22.
219. Kamper, S.J., *Sampling: Linking Evidence to Practice.* J Orthop Sports Phys Ther, 2020. **50**(12): p. 725-726.
220. Rysstad, T., et al., *Stratifying workers on sick leave due to musculoskeletal pain: translation, cross-cultural adaptation and construct validity of the Norwegian Keele STarT MSK tool.* Scand J Pain, 2022.
221. Munk Killingmo, R., et al., *Comparison of self-reported and public registered absenteeism among people on long-term sick leave due to musculoskeletal disorders: criterion validity of the iMTA Productivity Cost Questionnaire.* Eur J Health Econ, 2021. **22**(6): p. 865-872.
222. Werner, E.L. and C. Ihlebaek, *Primary care doctors' management of low back pain patients--ten years after.* Tidsskr Nor Laegeforen, 2012. **132**(21): p. 2388-90.
223. Patel, N.D., et al., *ACR Appropriateness Criteria Low Back Pain.* J Am Coll Radiol, 2016. **13**(9): p. 1069-78.
224. Henschke, N., et al., *Prognosis in patients with recent onset low back pain in Australian primary care: inception cohort study.* BMJ, 2008. **337**: p. a171.
225. Jensen, R.K., et al., *Lumbar spinal stenosis.* BMJ, 2021. **373**: p. n1581.
226. Kreiner, D.S., et al., *An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spinal stenosis (update).* Spine J, 2013. **13**(7): p. 734-43.
227. Matz, P.G., et al., *Guideline summary review: An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spondylolisthesis.* Spine J, 2016. **16**(3): p. 439-48.
228. Weiner, D.K., et al., *Low back pain in older adults: are we utilizing healthcare resources wisely?* Pain Med, 2006. **7**(2): p. 143-50.
229. Machado, G.C., et al., *Non-steroidal anti-inflammatory drugs (NSAIDs) for musculoskeletal pain.* BMJ, 2021. **372**: p. n104.
230. McCarberg, B.H., *NSAIDs in the older patient: balancing benefits and harms.* Pain Med, 2013. **14 Suppl 1**: p. S43-4.
231. Muller, A.E., et al., *Prescribed opioid analgesic use developments in three Nordic countries, 2006-2017.* Scand J Pain, 2019. **19**(2): p. 345-353.
232. Kim, H., et al., *Cultural Adaptation and Validation of the Korean Version of the iMTA Productivity Cost Questionnaire.* Healthcare (Basel), 2020. **8**(2).
233. Grovle, L., et al., *Poor agreement found between self-report and a public registry on duration of sickness absence.* J Clin Epidemiol, 2012. **65**(2): p. 212-8.
234. Voss, M., et al., *Comparisons of self-reported and register data on sickness absence among public employees in Sweden.* Occup Environ Med, 2008. **65**(1): p. 61-7.
235. Johns, G. and M. Miraglia, *The reliability, validity, and accuracy of self-reported absenteeism from work: a meta-analysis.* J Occup Health Psychol, 2015. **20**(1): p. 1-14.

Paper I

Killingmo RM, Storheim K, van der Windt DA,
Zolic-Karlsson Z, Vigdal ØN, Kretz L, Småstuen MC, Grotle M

Healthcare utilization and related costs among older people seeking primary care due to back pain: findings from the BACE-N cohort study

BMJ Open 2022 Jun 20;12(6):e057778.
DOI: 10.1136/bmjopen-2021-057778

BMJ Open

Healthcare utilization and related costs among older people seeking primary care due to back pain: findings from the BACE-N cohort study

Rikke Munk Killingmo ,[1] Kjersti Storheim,[1,2] Danielle van der Windt ,[3] Zinajda Zolic-Karlsson,[4] Ørjan Nesse Vigdal ,[1] Lise Kretz,[1] Milada Cvancarova Småstuen,[1] Margreth Grotle [1,2]

To cite: Killingmo RM, Storheim K, van der Windt D, *et al*. Healthcare utilization and related costs among older people seeking primary care due to back pain: findings from the BACE-N cohort study. *BMJ Open* 2022;**12**:e057778. doi:10.1136/bmjopen-2021-057778

► Prepublication history for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-057778).

Received 29 September 2021
Accepted 15 May 2022

[1]Department of Physiotherapy, Oslo Metropolitan University, Oslo, Norway
[2]Research and Communication Unit for Musculoskeletal Health (FORMI), Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway
[3]School of Medicine, Keele University, Keele, UK
[4]HTA and Reimbursement, Norwegian Medicines Agency, Oslo, Norway

Correspondence to
Rikke Munk Killingmo;
rikke@oslomet.no

ABSTRACT

Objectives To describe healthcare utilization and estimate associated costs during 1 year of follow-up among older people seeking primary care due to a new episode back pain and to describe healthcare utilization across patients with different risk profiles stratified using the StarT Back Screening Tool (SBST).

Design Prospective cohort study.

Participants and setting A total of 452 people aged ≥55 years seeking Norwegian primary care with a new episode of back pain were included.

Outcome measures The primary outcome of this study was total cost of healthcare utilization aggregated for 1 year of follow-up. Secondary outcomes included components of healthcare utilization aggregated for 1 year of follow-up. Healthcare utilization was self-reported and included: primary care consultations, medications, examinations, hospitalisation, rehabilitation stay, and operations. Costs were estimated based on unit costs collected from national pricelists. Healthcare utilization across patients with different SBST risk profiles was compared using Kruskal-Wallis test, post hoc Mann-Whitney U tests and Bonferroni adjustment.

Results In total, 438 patients were included in the analysis. Mean (BCa 95% CI) total cost per patient over 1 year was €825 (682-976). Median (BCa 95% CI) total cost was €364 (307-440). The largest cost category was primary care consultations, accounting for 56% of total costs. Imaging rate was 34%. The most commonly used medication was paracetamol (27%–35% of patients). Medium- and high-risk patients had a significantly higher degree of healthcare utilization compared with low-risk patients (p<0.030).

Conclusion This study estimated a 1 year mean and median cost of healthcare utilization of €825 and €364, respectively. Patients within the top 25th percentile accounted for 77% of all costs. Patients classified as medium risk and high risk had a significantly higher degree of healthcare utilization compared with patients classified as low risk.

Trial registration number ClinicalTrials.gov NCT04261309, results

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ The main strength of the present study is that it was conducted in line with the PROGnosis RESearch Strategy framework and preplanned with a published statistical analysis plan.
⇒ We used descriptive statistics to conduct an overall prognosis study and provide evidence to inform quality improvement in primary care management of back pain.
⇒ The main limitation with this study is that we had missing data (18.4% to 26.0%) on variables used to estimate the outcome variables and had to manually replace missing values.
⇒ Due to differences in primary care organisation between countries, readers are advised to exercise caution with generalisation of the results to other healthcare systems.

INTRODUCTION

The burden of back pain has been growing along with an increasing and ageing population.[1–4] In recent years, back pain has become the leading cause of disability globally[4 5] and an extensive burden to our healthcare systems.[1 6–8] According to a recent systematic review, the prevalence rate of healthcare utilization for back pain ranges from 28% to 92%,[9] and patients with back pain have previously been shown to consume close to two times as much healthcare as the general population.[10] Physiotherapists, chiropractors and general practitioners (GP) are healthcare providers commonly engaged in the management of back pain.[9] Back pain is one of the most prevalent complaints encountered in primary care.[3 8 11] In Norway, a former study has shown that back pain accounts for as many as 27, 82 and 10% of all consultations to physiotherapists, chiropractors and GPs, respectively.[12]

Updated international clinical guidelines provide, more or less, consistent recommendations for how to assess and treat patients with back pain.[13–16] A key recommendation

is to adopt a stratified healthcare approach, guided by the patients response to care or the results of risk prediction tools (such as the StarT Back Screening Tool (SBST),[7 14 17 18] which has been shown to be a cost-effective strategy in primary care.[19] As targeting resources to those most likely to benefit might allow an improvement in patient outcomes while reducing avoidable costs and the burden on healthcare systems.[14 18–20]

Although these guidelines are well established and health providers report being aware of them, concerns about substantial gaps between guidelines and practice have been highlighted. Problems include both underuse of high-value care (eg, education, advice to remain active and exercise), overuse of low-value care (eg, pharmacological treatment as first-line treatment and high imaging rates), and thereby misuse of limited healthcare resources.[1 2 13 14] The extent to which this concern also applies to older people seeking primary care due to back pain is unknown. Historically older people have been underrepresented in back pain research,[21–23] though in recent years, cohort studies have been designed to specifically investigate the course and prognosis of back pain in older people.[24 25] To improve use of scarce resources and thus reduce the burden on our healthcare systems, researchers have highlighted the importance of monitoring and understanding healthcare utilisation and costs related to back pain.[2 14]

Therefore, the primary aim of this study was to describe healthcare utilization and estimate associated costs during 1 year of follow-up among older people seeking primary care due to a new episode of back pain. The secondary aim was to describe healthcare utilisation across patients with different risk profiles stratified according to the SBST.

METHOD

This study is designed and performed in accordance with the PROGnosis RESearch Strategy (PROGRESS) framework[26] and is considered part of overall prognosis research. In line with recommendations from the PROGRESS framework,[26] a study protocol including a statistical analysis plan has been published (ClinicalTrials.gov Identifier: NCT04261309).[27]

Design and setting

This study presents data from the Back Complaints in the Elderly—Norway study (BACE-N), a prospective observational cohort study with 1 year of follow-up within a Norwegian primary care setting. The BACE-N is part of the international BACE consortium.[24]

Participants and recruitment procedure

Eligible participants were people 55 years of age or older seeking primary care (physiotherapist, chiropractor or GP) with a new episode of back pain (preceded by 6 months without visiting a primary care provider for similar complaints). Patients were excluded if they had difficulties completing the questionnaires (eg, unable to speak, read or write in Norwegian) or if they had difficulties completing the physical examination (eg, are wheelchair-bound). Patients were recruited from physiotherapists, chiropractors and GPs working in Norwegian primary care between April 2015 and February 2020. Patients who met the eligibility criteria and completed the consent to participate were included in the study.

Data collection, outcome, screening tool and other variables

At baseline, all patients responded to a comprehensive questionnaire and went through a standardised physical examination conducted by local research assistants at test stations established within each recruiting area. Follow-up questionnaires were sent at 3, 6 and 12 months after inclusion. All questionnaires were preferably completed electronically, but paper versions were available for patients not familiar with electronic data collection. Within this study, only data from questionnaires were used.

Outcome variables

The primary outcome of this study was total cost of healthcare utilization aggregated for 1 year of follow-up. Secondary outcomes included components of healthcare utilization aggregated for 1 year of follow-up.

Healthcare utilization was self-reported and included: consultations to healthcare professionals (type and frequency), use of back medication (both prescription and over-the-counter, type and frequency), number of diagnostic examinations (type and frequency), number of days of hospitalisation and/or rehabilitation stay and back operations. Consultations to healthcare professionals and use of back medication were reported with a 3-month recall period at each timepoint of follow-up. Number of diagnostic examinations and days of hospitalisation and/or rehabilitation stay were reported with a 3-month recall period at 3-month and 6-month follow-up and a 6-month recall period at 12-month follow-up. Back operations were reported with a 12-month recall period at 12-month follow-up. Total costs of healthcare utilization per patient were estimated by multiplying frequency of use by unit costs collected from national pricelists (see table 1). Non-healthcare costs related to provision of healthcare (as transportation) were not estimated. Costs related to back medication were estimated based on medication type (not exact medication name) and frequency of use (data on dosage were not available).

Screening tool

The SBST[17] was used to classify included patients into low, medium or high risk of poor disability outcome. The SBST is a brief 9-item tool designed to screen primary care patients with low back pain for prognostic indicators that are relevant to initial decision-making. The tool is summed to produce an overall score from 0 to 9 and a psychological subscale score from 0 to 5. Patients with an overall score between 0 and 3 are classified as low risk. Patients with an overall score of minimum 4 and a

Table 1 Cost categories, units, unit price, all numbers in euros (€) and Norwegian kroner (NOK) for 2020

Cost categories	Unit	Unit price (€)	Unit price (NOK)	Reference (source)
Primary care				
General practitioner	Per visit	43.1	431	The Norwegian Medical Association, estimated average
Physiotherapist	Per visit	47.2	472	The Norwegian Physiotherapy Association, estimated average
Chiropractor	Per visit	55.0	550	Private price lists, estimated average
Manuel therapist	Per visit	74.2	742	The Norwegian Physiotherapy Association, estimated average
Naprapath	Per visit	64.0	640	Private price lists, estimated average
Osteopath	Per visit	65.0	650	Private price lists, estimated average
Psychologist	Per visit	110.0	1100	The Norwegian Psychological Association, estimated average
Other therapists	Per visit	75.0	750	Private price lists, estimated average
Medication				
Paracetamol	Per daily defined dose	0.5	5	NoMA price list, estimated average
NSAID	Per daily defined dose	1.2	12	NoMA price list, estimated average
Muscle relaxant	Per daily defined dose	0.7	7	NoMA price list, estimated average
Sleep medication	Per daily defined dose	0.2	2	NoMA price list, estimated average
Cortisone	Per daily defined dose	0.4	4	NoMA price list, estimated average
Opioid	Per daily defined dose	0.9	9	NoMA price list, estimated average
Examination				
Blood sample	Per examination	20.4	204	The Norwegian Medical Association, estimated average
X-ray	Per examination	119.0	1190	Unilabs price list, estimated average
MRI	Per examination	269.0	2690	Unilabs price list, estimated average
CT	Per examination	189.0	1890	Unilabs price list, estimated average
Secondary care				
Back operation	Per operation	5220.0	52200	DRG2150
Hospitalisation (non-operation)	Per day	1880.0	18800	The Norwegian Directorate of Health, SAMDATA
Rehabilitation stay	Per day	315.0	3150	UniCare pricelist, estimated average

NoMA, Norwegian Medicines Agency; NSAID, non-steroidal anti-inflammatory drug.

subscale score of maximum 3 are classified as medium risk. Patients with an overall score of minimum 4 and a subscale score of 4 or 5 are classified as high risk.

The SBST has been recommended in guidelines to enable stratified care for patients with low back pain.[14 18] Simpler and less intensive support should be considered for people who are likely to improve quickly and have a good outcome. More complex and intensive support should be considered for people with higher risk of a poor outcome. The SBST was translated into Norwegian by Storheim and Grotle in 2012 and has shown to have an acceptable accuracy in predicting persistent disabling back pain.[17 28–31]

Other variables

Overall prognosis may vary depending on context (time, place, healthcare setting) and characteristics of the study population. In line with the PROGRESS framework and recommendations for overall prognosis studies,[26] descriptive variables were based on previous scientific literature and included the following variables measured at baseline:

- Sex[32–35] (female/male).
- Age[32–35] (years).
- Educational level[36 37] measured as the highest education completed and categorised into low (elementary and high school level) or high (university level).
- First healthcare provider[38] (physiotherapist, chiropractor or GP).
- Pain severity[33 34 39–42] measured by the Numeric Rating Scale (range 0–10, higher score indicate higher pain severity).[43]
- Pain duration[39] measured by the question 'how many days have you had your current back pain?'
- Pain history[40] measured by the question 'have you had back pain before?'
- Radiating pain below the knee[41] measured by the question 'did your back pain radiate to your legs last week? If yes, how far down did the pain radiate?'
- Disability[33 34 37 39–41] measured by the Roland-Morris Disability questionnaire (range 0–24, higher score indicates higher degree of back-related disability).[44]
- Comorbidity[42 45 46] measured by the Self-Administered Comorbidity Questionnaire (13 predefined comorbidities and two optional comorbidities. Item number 12 (back pain) was replaced with a third optional comorbidity).[47]
- Health-related quality of life[34 42] measured by the Short-Form Health Survey 36-item physical and mental summary score (range 0–100, higher score indicate better health-related quality of life).[48]
- Emotional well-being[37 39 41 45 49] measured by the Centre for Epidemiological Studies-Depression questionnaire (range 0–60, higher score indicates more signs of depression).[50]
- Kinesiophobia[41 49] measured by the Fear Avoidance Beliefs Questionnaire—Physical Activity subscale (range 0–24, higher score indicates higher levels of kinesiophobia).[51]
- Red flags (cancer, first episode of back pain, constant pain, unexplained weight loss, systematically unwell, fever, urinary retention or loss of bladder control, age ≥75 years, trauma cause of back pain, osteoporosis, cortisone use and severe morning stiffness).[52 53]
- Total costs related to healthcare utilization prior to inclusion measured in the period from baseline to 6 weeks retrospectively. Healthcare utilization prior to inclusion was self-reported and included: primary care consultations, use of back medication and number of diagnostic examinations. Total cost of healthcare utilization was estimated by multiplying frequency of use by unit costs collected from national pricelists (see table 1).

In addition, included patients were described with respect to ethnicity and pain location.

Analyses

The statistical analysis plan for this study was informed by recommendations from the PROGRESS framework.[26] All analyses are outlined in the statistical analysis plan published a priori[27] and performed using the IBM SPSS V.26 (IBM Corporation, Armonk, New York). P values <0.05 were considered statistically significant. All statistical tests were two sided.

Study flow

The flow of participants through the study was reported according to the STROBE guidelines[54] with a flowchart. Reasons for dropout were provided where known. Baseline differences between responders and non-responders at 12-monthfollow-up were evaluated. Mann-Whitney U test was used for continuous variables. Pearson χ^2 and Fisher's exact test (if <5 cases in one cell) were used for categorical variables.

Missing data

Missing value pattern was visually explored, and missingness at random was assumed. Also, we found evidence against the hypothesis that values were not missing completely at random (Little's test, p>0.05). Missing baseline data were handled by multiple imputation within the BACE-N. Five multiple imputation data sets with 10 iterations were created using regression estimation. We did not impute missing outcome values, as the imputation model had poor predictive performance and caused a clear trend of values being overestimated. Instead, missing values on variables used to calculate the outcome scores were imputed with: (1) each patient's individual average of observed values for the variables: consultations to healthcare professionals and medication use, (2) a value of zero costs for the variables: diagnostic examinations, hospitalisation, rehabilitation stay and back operations.

Healthcare utilization and cost estimation

Type and frequency of use of different healthcare resources were calculated for each of the follow-up

periods. All costs were presented in euros (€) 2020 and estimated with both mean and median values with 95% CI, using bias-corrected and accelerated (BCa) bootstrapping for each follow-up period and the whole year. The BCa was conducted with a bootstrap sample size of 1000. Cost data are commonly skewed, thus both mean and median values were presented to support the result interpretation. Values in Norwegian kroner (NOK) were recalculated to euros using the exchange rate from February 2020 (1€=NOK 10).

Healthcare utilization across patients with different risk profiles

Type and frequency of use of different healthcare resources were described for the 1-year follow-up, for the following subgroups: (1) low, (2) medium and (3) high risk of persistent disabling back pain according to the SBST. The Kruskal-Wallis test including post hoc Mann-Whitney U tests with Bonferroni adjustment were conducted to determine between-group differences with regards to number of primary care consultations, number of patients using back medication, number of patients receiving imaging (X-ray, MRI, CT) and number of patients receiving secondary care (back operation, hospitalisation, rehabilitation stay). The Bonferroni adjustment was applied by multiplying raw P values by the number of tests conducted (0.05×3).

Sensitivity analysis

To test credibility of the manual imputation on missing values used to calculate the outcome scores and total cost calculations related to the primary analyses, two sensitivity analyses were performed; (1) complete case analysis without adjustment for missing data and (2) without outliers. Outliers were identified with simple scatterplots by visual inspection and defined as patients with remarkably high total costs at each time period; 5 patients with costs ≥ €2433 at 0–3 months, 5 patients with costs ≥€6025 at >3–6 months, 8 patients with costs ≥€3518 at >9–12 months and 11 patients with costs ≥€8004 at 0–12 months. All outliers were patients with healthcare utilisation within secondary care, primarily hospitalisation and operations.

Sample size

This study contains secondary analyses embedded in the BACE-N. Details on sample size calculation are provided in the BACE-N protocol.[27] We considered a sample size of 450 participants within the BACE-N to be sufficient to describe healthcare utilisation and estimate associated costs.[55]

Patient and public involvement

Patient representatives were part of the scientific board of the study and involved in designing and establishing BACE-N. Results will be disseminated to the recruiting primary care providers and the participating patients in an annual newsletter.

RESULTS

A total of 452 patients were included in this study. Table 2 shows patient characteristics and clinical status at baseline, along with the proportion with missing data per variable. Flow of patients through the study is shown in figure 1. Fourteen patients (3%) were dropouts at 12-month follow-up and were, thus, removed from the analyses. There was a larger proportion of women (55 vs 42%) among the responders as compared with non-responders. Otherwise, there were no differences between responders and non-responders.

Missing data ranged from 0.0% to 16.8% for included baseline variables and 18.4% to 26.0% for healthcare variables used to calculate the outcome values. Total missingness was 4.9% and 23.3% for all baseline and follow-up values, respectively.

Healthcare utilization and cost estimation

Table 3 shows healthcare utilization throughout 1 year of follow-up. Table 4 shows costs related to healthcare utilization for each follow-up period and aggregated for 1 year of follow-up. Almost all included patients (87%) had costs related to healthcare utilization during the 1 year of follow-up. Nevertheless, the distribution of costs was highly skewed to the left, indicating that most of costs emerged from a minority of the patients. Patients within the top 5th, 10th and 25th percentile accounted for, respectively, 43%, 55% and 77% of total costs within the sample. The mean (BCa 95% CI) and median (BCa 95% CI) total cost per patient for 1 year of follow-up were estimated at €825 (682–976) and €364 (307–440), respectively. The largest cost category was primary care consultations, accounting for 56% of total costs. The remaining cost categories; back medication, examination, hospitalisation, rehabilitation stay and back operation accounted for 6, 8, 16, 3 and 11% of total costs, respectively.

The sensitivity analyses showed no substantial change in point estimates when comparing complete case analysis and analysis without outliers to the main analysis. The complete case analysis provided an estimated mean (BCa 95% CI) and median (BCa 95% CI) of total cost per patient for 1 year of follow-up at €873 (670–1116) and €343 (280–463), respectively. Furthermore, the analysis without outliers provided an estimated mean (BCa 95% CI) and median (BCa 95% CI) of total cost per patient for 1 year of follow-up at €573 (505–635) and €340 (277–416), respectively.

Healthcare utilization across patients with different risk profiles

Table 5 shows healthcare utilization throughout 1 year of follow-up across patients with different risk profiles according to the SBST. The SBST classified 289 patients (66%) as low, 120 (27%) as medium and 29 (7%) as high risk of persistent disabling back pain, respectively. Healthcare utilization increased with increasing degree of risk of persistent disabling back pain according to formal testing with the Kruskal-Wallis test, including post

Table 2 Patient characteristics and clinical status at baseline*

	All participants (n=452)	Missing, n (%)	Stratified risk profile†		
			Low (n=297)	Medium (n=125)	High (n=30)
Female	235 (52)	0 (0)	137 (46)	78 (62)	20 (67)
Age in years	66 (59–72)	0 (0)	66 (59–72)	65 (58–73)	70 (65–77)
Educational level high	199 (44)	20 (4)	140 (47)	48 (39)	10 (33)
Ethnicity Norwegian	430 (95)	0 (0)	287 (97)	116 (93)	27 (90)
First healthcare provider					
General practitioner	127 (28)	0 (0)	51 (17)	26 (21)	7 (23)
Physiotherapist	130 (29)	0 (0)	107 (36)	41 (33)	12 (40)
Chiropractor	195 (43)	0 (0)	139 (47)	58 (46)	11 (37)
Pain location					
Thoracic	61 (14)	11 (2)	37 (12)	21 (17)	3 (10)
Lumbar/sacral	414 (92)	11 (2)	273 (92)	112 (90)	29 (97)
Radiating pain below the knee	141 (31)	0 (0)	66 (22)	63 (50)	12 (40)
Pain severity average last week (NRS, 0–10)	5 (4–7)	31 (7)	5 (3–7)	7 (5–8)	7 (5–8)
Pain duration					
<6 weeks	297 (66)	76 (17)	194 (65)	89 (71)	14 (47)
6 weeks to 3 months	59 (13)	76 (17)	37 (13)	16 (13)	6 (20)
>3 months	96 (21)	76 (17)	66 (22)	20 (16)	10 (33)
Previous episodes of back pain	426 (94)	29 (6)	279 (94)	120 (96)	27 (90)
Disability (RMDQ 0–24)	9 (4–13)	45 (10)	6 (3–10)	13 (10–16)	17 (13–19)
Comorbidity (SCQ, 0–15)	1 (1–2)	18 (4)	1 (0–2)	2 (1–3)	2 (2–3)
Health-related QOL (SF36, 0–100)					
Physical component	42 (36–47)	41 (9)	45 (39–50)	37 (33–43)	33 (30–39)
Mental component	55 (47–60)	41 (9)	57 (51–61)	51 (43–56)	38 (29–48)
Emotional well-being (CES-D, 0–60)	8 (4–15)	57 (13)	6 (3–11)	12 (8–18)	18 (15–29)
Kinesiophobia (FABQ-PA, 0–24)	10 (5–13)	18 (4)	10 (5–15)	15 (10–19)	19 (15–22)
Numbers of red flags (0–12)	1 (0–2)	50 (11)	1 (0–1)	1 (1–2)	3 (1–4)
Healthcare utilization prior to inclusion					
Primary care consultation last 6 weeks					
General practitioner	83 (18)	21 (5)	47 (16)	24 (19)	12 (40)
Physiotherapist	129 (29)	21 (5)	87 (29)	32 (26)	10 (33)
Chiropractor	188 (42)	21 (5)	123 (41)	56 (45)	9 (30)
Manual therapist	19 (4)	21 (5)	13 (4)	6 (5)	0 (0)
Naprapath	15 (3)	21 (5)	8 (3)	5 (4)	1 (3)
Osteopath	3 (1)	21 (5)	2 (0.7)	0 (0)	1 (3)
Psychologist	2 (0.4)	21 (5)	1 (0.3)	1 (0.8)	0 (0)
Other therapists	7 (2)	21 (5)	4 (1)	2 (2)	1 (3)

Continued

Table 2 Continued

	All participants (n=452)	Missing, n (%)	Stratified risk profile†		
			Low (n=297)	Medium (n=125)	High (n=30)
Use of medication	189 (42)	38 (8)	94 (32)	71 (57)	23 (77)
Diagnostic examination last 6 months					
Blood sample	12 (3)	24 (5)	7 (2)	0 (0)	5 (17)
X-ray	26 (6)	24 (5)	12 (4)	7 (6)	7 (23)
MRI	53 (12)	24 (5)	30 (10)	15 (12)	8 (27)
CT	8 (2)	24 (5)	6 (2)	1 (0.8)	1 (3)
Previous hospitalisation	54 (12)	21 (5)	24 (8)	18 (14)	12 (40)
Previous rehabilitation stay	18 (4)	25 (6)	7 (2)	7 (6)	4 (13)

All values are presented by number (percentage of total) or median (IQR).
*The presented characteristics are pooled estimates based on the multiple imputation procedures.
†According to the StarT Back Screening Tool.
CES-D, Center for Epidemiological Studies-Depression; FABQ-PA, Fear-Avoidance Beliefs Questionnaire-Physical Activity subscale; NRS, numeric rating scale; RMDQ, Roland-Morris Disability Questionnaire; SCQ, Self-administered Comorbidity Questionnaire; SF-36, Short Form Health Survey 36 Item.

hoc Mann-Whitney U tests: low-risk patients had fewer primary care consultations (p<0.001), used less frequently back medication (p<0.001) and received less frequently imaging (p<0.003) and secondary care (p<0.030), compared with medium-risk patients. Moreover, low-risk patients had fewer primary care consultations (p<0.001),

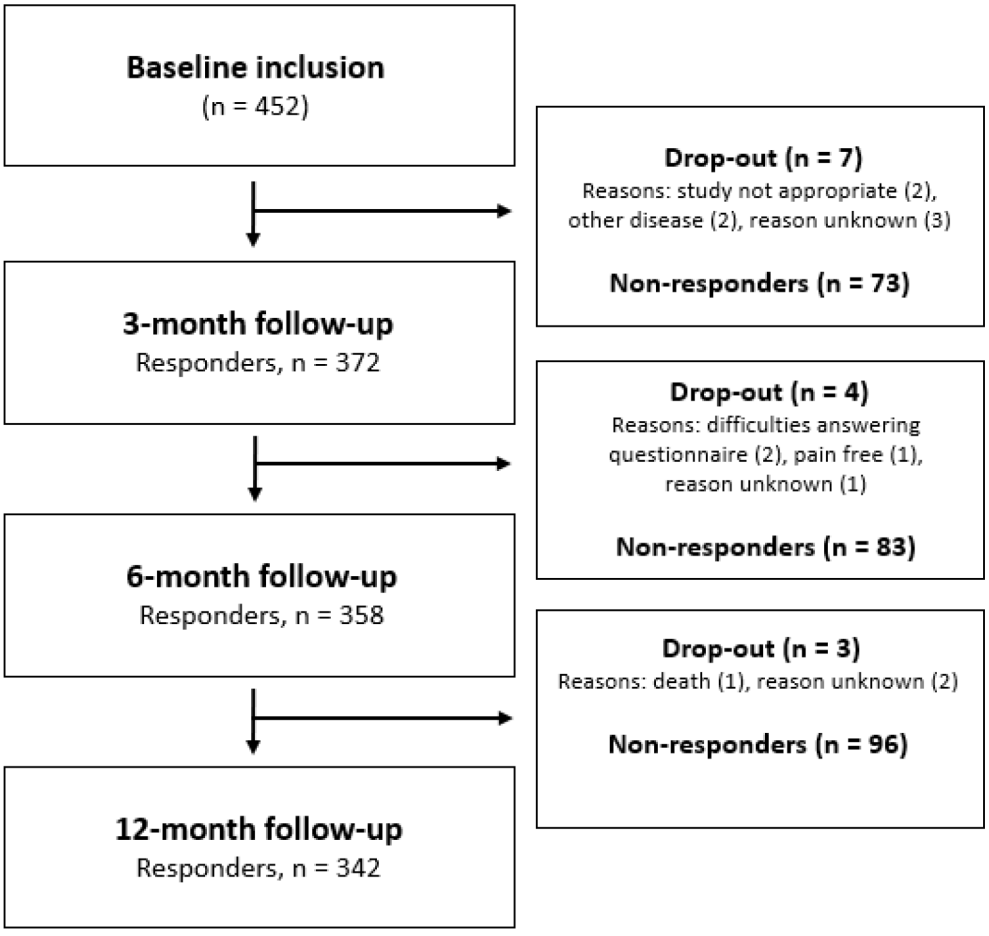

Figure 1 Participant flowchart.

Table 3 Healthcare utilization throughout 1 year of follow-up (n=438)

	0–3 months		>3–6 months		>9–12 months	
		Missing, n (%)		Missing, n (%)		Missing, n (%)
Primary care						
Primary care consultation, N (%)		79 (18)		87 (20)		108 (24)
General practitioner	44 (12)		30 (9)		22 (7)	
Physiotherapist	119 (33)		70 (20)		48 (15)	
Chiropractor	124 (35)		76 (22)		50 (15)	
Manual therapist	22 (6)		5 (1)		7 (2)	
Naprapath	6 (2)		11 (3)		6 (2)	
Osteopath	2 (0.6)		1 (0.3)		3 (1)	
Psychologist	0 (0)		1 (0.3)		1 (0.3)	
Other therapists	10 (3)		12 (3)		7 (2)	
No primary care consultations	93 (26)		179 (51)		212 (64)	
Numbers of consultations, median (IQR)*						
General practitioner	1 (1–2)	0 (0)	1 (1–2)	0 (0)	1 (1–3)	0 (0)
Physiotherapist	4 (2–8)	0 (0)	4 (2–10)	2 (3)	5 (1–9)	0 (0)
Chiropractor	4 (2–6)	0 (0)	2 (1–4)	4 (5)	3 (1–5)	0 (0)
Manual therapist	3 (1–5)	0 (0)	3 (2–14)	0 (0)	1 (1–4)	0 (0)
Naprapath	3 (1–5)	0 (0)	4 (2–6)	0 (0)	3 (1–4)	0 (0)
Osteopath	3 (2-)	0 (0)	2 (2–2)	0 (0)	10 (2-)	0 (0)
Psychologist	–	–	1 (1–1)	0 (0)	7 (7–7)	0 (0)
Other consultations	4 (1–6)	0 (0)	1 (1–8)	0 (0)	4 (2–8)	1 (14)
Medication						
Use of back medication, N (%)		80 (18)		96 (22)		114 (26)
Paracetamol	124 (35)		91 (27)		86 (27)	
NSAID	86 (24)		75 (22)		64 (20)	
Muscle relaxants	6 (2)		4 (1)		3 (1)	
Sleep medication	22 (6)		22 (6)		13 (4)	
Cortisone	5 (1)		9 (3)		4 (1)	
Opioid	5 (1)		5 (2)		3 (1)	
No use of back medication	197 (55)		213 (62)		213 (66)	
Frequency of use paracetamol, N (%)**						
Daily	46 (37)	0 (0)	32 (35)	0 (0)	30 (35)	0 (0)
Weekly	35 (28)	0 (0)	30 (33)	0 (0)	28 (33)	0 (0)
Monthly or less	43 (35)	0 (0)	29 (32)	0 (0)	28 (32)	0 (0)
Frequency of use NSAID, N (%)†						
Daily	22 (26)	0 (0)	16 (21)	0 (0)	17 (26)	0 (0)
Weekly	14 (16)	0 (0)	25 (33)	0 (0)	19 (30)	0 (0)
Monthly or less	50 (58)	0 (0)	34 (46)	0 (0)	28 (44)	0 (0)

Continued

Table 3 Continued

	0–3 months		>3–6 months		>9–12 months	
		Missing, n (%)		Missing, n (%)		Missing, n (%)
Frequency of use opioid, N (%)†						
Daily	3 (60)	0 (0)	4 (80)	0 (0)	2 (67)	0 (0)
Weekly	1 (20)	0 (0)	–	–	–	–
Monthly or less	1 (20)	0 (0)	1 (20)	0 (0)	1 (33)	0 (0)
Examination						
Diagnostic examination, N (%)		79 (18)		86 (20)		106 (24)
Blood sample	9 (3)		5 (1)		6 (2)	
X-ray	12 (3)		8 (2)		16 (5)	
MRI	37 (10)		17 (5)		20 (6)	
CT	4 (1)		2 (1)		2 (1)	
No diagnostic examination	281 (77)		316 (89)		289 (87)	
Secondary care						
Back operation, N (%)	–	–	–	–	7 (2)	103 (24)
Hospitalisation, N (%)	5 (1)	75 (17)	6 (2)	84 (19)	2 (1)	104 (24)
Duration of stay in days, median (range)	1 (1–2)	0 (0)	3 (2–5)	1 (17)	2.5 (2-)	0 (0)
Rehabilitation stay, N (%)	0 (0)	73 (17)	1 (0.3)	84 (19)	1 (0.3)	104 (24)
Duration of stay in days, median (range)	–	–	20 (20–20)	0 (0)	7 (7–7)	0 (0)

Cells marked with a dash (-) indicate that the variable was not reported.
*Numbers of consultations are calculated on the basis of patients who have reported primary care consultations.
†Frequency of back medication use is calculated on the basis of patients who have reported back medication use.
NSAID, non-steriodal anti-inflammatory drug.

used less frequently back medication (p<0.001) and received less frequently imaging (p<0.015), compared with high-risk patients. No differences were revealed between medium-risk and high-risk patients.

DISCUSSION

The present study describes the prevalence and associated costs of healthcare utilization among older people seeking primary care due to a new episode of back pain. The mean and median total cost per patient during the 1 year of follow-up was €825 and €364, respectively. The largest cost category was primary care consultations. Patients within the top 25th percentile accounted for 77% of all costs. Patients with medium-risk and high-risk of poor disability had a significantly higher degree of healthcare utilization compared with patients with low risk.

Direct comparability of this study with other studies is limited. To the best of our knowledge, no similar study has been conducted among a sample of exclusively older people with back pain or within the Norwegian healthcare system.[56] Furthermore, there is a widespread heterogeneity in the methodologies used among back pain cost of illness studies.[56 57] Nevertheless, several of our findings are generally in accordance with previous research on primarily middle-aged patients with back pain. The majority of cost of illness studies recruiting participants from primary care have estimated in 2020 euros a 1-year mean total direct cost related to back pain per patient ranging from €1.000 to €2.000.[41 56 58 59] Furthermore, several studies have found that primary care consultations are frequently used and a large cost category among patients with back pain,[8 33 56–62] and that the majority of healthcare utilization and related costs stem from a relatively small group of patients.[61 63 64] In the present study, descriptive statistics indicated a gradual decrease in costs related to primary care and a gradual increase in costs related to secondary care during the 1 year of follow-up. Yet, that result should be interpreted with caution, especially for costs related to secondary care where the mean values deviated to a fairly large extent from the median values, hence indicating that the increase is largely due to a few individuals with (remarkably) high costs.

Table 4 Cost (€) due to healthcare utilization from 0 to 3 months, >3–6 months, >9–12 months and 0–12 months* (n=438)

	Patients with zero costs, n (%)	0–3 months		>3–6 months		>9–12 months		0–12 months*	
		Mean (95% CI)†	Median (95% CI)†	Mean (95% CI)†	Median (95% CI)†	Mean (95% CI)†	Median (95% CI)†	Mean (95% CI)†	Median (95% CI)†
Primary care	83 (21)	199 (178 to 222)	116 (94 to 154)	138 (118 to 161)	43 (24 to 47)	120 (98 to 145)	0 (0 to 0)	458 (404 to 516)	242 (192 to 330)
Medication	176 (44)	19 (15 to 23)	0 (0 to 0.4)	17 (14 to 21)	0 (0 to 0)	16 (13 to 20)	0 (0 to 0)	52 (43 to 61)	3 (1 to 7)
Examination	308 (77)	31 (23 to 39)	0 (0 to 0)	15 (9 to 21)	0 (0 to 0)	19 (13 to 26)	0 (0 to 0)	65 (50 to 81)	0 (0 to 0)
Secondary care	390 (97)	33 (9 to 61)	0 (0 to 0)	90 (28 to 162)	0 (0 to 0)	120 (50 to 216)	0 (0 to 0)	243 (116 to 388)	0 (0 to 0)
Total	52 (13)	281 (244 to 322)	165 (165 to 165)	261 (189 to 346)	55 (46 to 55)	276 (195 to 370)	44 (23 to 46)	825 (682 to 976)	364 (307 to 440)

*Cost due to healthcare utilisation for the entire follow-up period is calculated on the basis for the three follow-up periods.
†Bias-corrected and accelerated bootstrapping (1000 simulations).

Table 5 Healthcare utilization throughout 1 year of follow-up, across patients with different risk profile according to the StarT Back Screening tool (n=438)*

	Stratified risk profile		
	Low (n=289)	Medium (n=120)	High (n=29)
Primary care			
Primary care consultation, N (%)	205 (76)	94 (86)	21 (88)
Numbers of consultations, median (IQR)†	5 (3–11)	12 (6–19)	15 (8–22)
Medication			
Use of back medication, N (%)	128 (48)	77 (71)	21 (91)
Paracetamol	95 (35)	68 (63)	18 (78)
NSAID	88 (33)	39 (36)	10 (44)
Muscle relaxants	1 (0.4)	7 (7)	3 (13)
Sleep medication	14 (5)	11 (10)	8 (35)
Cortisone	4 (2)	4 (4)	5 (22)
Opioid	4 (2)	4 (4)	2 (9)
Examination			
Diagnostic examination, N (%)	73 (27)	45 (42)	12 (50)
Blood sample	14 (5)	3 (3)	2 (9)
X-ray	15 (6)	10 (9)	5 (22)
MRI	30 (11)	24 (22)	6 (26)
CT	3 (1)	4 (4)	1 (4)
Secondary care			
Back operation, N (%)	4 (2)	1 (1)	2 (11)
Hospitalisation, N (%)	4 (2)	6 (6)	2 (9)
Rehabilitation stay, N (%)	0 (0)	2 (2)	0 (0)

Valid percentages are given and have been rounded off.
*Healthcare utilization throughout 1 year of follow-up is calculated on the basis for the three follow-up periods.
†Number of consultations is calculated on the basis of patients who have reported primary care consultations.
NSAID, non-steriodal anti-anflammatory drug.

In the present study, we revealed an imaging rate of 34% during the 1 year of follow-up, including the time period from baseline to 6 months retrospectively. Comparably, Werner and Ihlebæk[65] showed that 39% of patients with low back pain in 2011 were referred for imaging by GPs in Norway. Likewise, in a recent systematic review of healthcare provided for patients with low back pain, Kamper *et al*[13] reported that around one in four was referred for imaging in family practice. Updated clinical guidelines recommend that imaging should not be routinely used, but rather reserved for patients for whom the result is likely to change management.[14 18 66] Also, evidence suggests that

prevalence of serious pathology as cause of back pain, for which imaging is indicated, in primary care is ≤6%.[1 52 67 68] In that context, a rate of 34% seems to indicate an overuse of imaging.[66 69]

Our findings regarding medication use are slightly different from previous research. In our study, paracetamol (27%–35%) followed by NSAIDs (20%–24%) were most commonly used, whereas only a small proportion of patients used opioids (1%–2%). Estimates provided by Kamper *et al*[13] have suggested that around 20% of low back pain patients within family practice are recommended paracetamol, 35%–40% NSAIDs and up to 30% opioids. Differences in paracetamol use might be explained by the fact that most studies do not include over-the-counter medication, thus use of paracetamol is probably underrepresented within the review by Kamper *et al*.[13] Differences in NSAIDs use might be explained by the fact that our sample consists of exclusively older people who often have a higher risk of NSAID-related side effects.[70 71] Differences in opioid use might be explained by the fact that Norway has strict opioid prescription regulations.[72] Updated clinical guidelines recommend pharmacological treatment as an adjunctive option in case of an inadequate response to first-line treatment.[14 18] NSAIDs should be first-line pharmacological treatment, taking into account possible side effects. Opioids should be used only in carefully selected patients. Paracetamol is not recommended. In that context, it appears that opioid use within this study might be in line with clinical guidelines, as opposed to paracetamol use.

Low-risk patients had a significantly lower degree of healthcare utilization compared with medium-risk and high-risk patients. We revealed no difference in healthcare utilization between medium-risk and high-risk patients. Yet, that result should be interpreted with caution due to a small sample size within the high-risk subgroup, thus risk of low statistical power. Updated clinical guidelines recommend a stratified healthcare approach.[7 14 18] In that context, it is promising that low-risk patients have a lower degree of healthcare utilization compared with medium-risk and high-risk patients.

The main limitation with this study is that we had missing data on variables used to estimate the outcome variables and had to manually replace missing values. It is well known that healthcare utilization is prone to missing data.[73–75] Also, that missing values should be replaced in order to make use of all reported data.[73 74] Unfortunately, due to poor predictive performance, multiple imputation could not be used in this study. We, therefore, chose a frequently used, though not optimal, method for replacing missing values and have been transparent in our reporting. A second limitation is the fact that we expect to have somewhat underestimated total healthcare utilization and related costs. Self-reports tend to underestimate the true value of healthcare utilization due to potential recall bias.[76–79] Furthermore, we lack data on primary care consultations and medication use between 6 and 9 months. A third limitation is the lack of data on eligible participants that declined to participate or for other reasons were not invited. Due to limited resources and practical reasons related to recruitment from a broad network of clinicians, it was not possible to record information on all eligible participants during the data collection period. To compensate for this limitation and assess the representativeness of the BACE-N sample, it has previously been compared on key sociodemographic variables with a subsample from a longitudinal population study: 'The Norwegian study on life course, ageing and generation (NORLAG)'.[80 81] The subsample (NORLAG MSK) is expected to be a representative sample of people aged ≥55 years with musculoskeletal complaints. Characteristics of the two samples were largely comparable, though the BACE-N sample has more men, and more with higher education levels. Previous studies have shown that women[33 34 40] are more likely to seek care for back pain as are people with lower education levels.[33 36 37] In that context, it is likely to assume that the amount of healthcare utilization presented in this study is somewhat underestimated. Furthermore, the BACE-N sample is largely comparable to younger Norwegian back pain cohorts[82 83] and to the BACE cohort from the Netherlands.[84] A fourth potential limitation, which might have affected the representativeness of the BACE-N sample, is that we used an age cut point of ≥55 years to define a population of older people. Commonly, older people are defined as those aged 60 or 65 years or older,[85] whereas in BACE-N, only 74% and 58% of patients were ≥60 and 65 years at baseline, respectively. An age cut point of ≥55 years within the BACE-N was determined based on the standardised methodology of the BACE consortium,[24] as this would allow comparisons across different countries. Within the BACE consortium, the decision of the age cut point was based on an age cut point (of ≥55 years), which was used in a large population cohort study of older people in the Netherlands (The Rotterdam Study).[86] Finally, a fifth potential limitation is that we conducted this study from a health system perspective, thus, indirect costs related to productivity loss were not estimated. Indirect costs are expected to have a strong impact on total costs related to back pain.[55] Therefore, this should be taken into account when interpreting the results.

The main strength of the present study is that it was conducted in line with the PROGRESS framework[26] and preplanned with a published statistical analysis plan. Also, it is the first study to estimate healthcare utilization and related cost among a sample of exclusively older people with back pain. Mapping healthcare utilization is vital to improve use of scarce healthcare resources and reduce the burden on our healthcare systems, where possible and appropriate.[2 14] This study addressed potential gaps between guidelines and practice; the use of paracetamol and imaging seems to be important areas for quality improvement in primary care management of older people with back pain.

Conclusion

In conclusion, this study estimated a 12-month mean and median cost of healthcare utilization of €825 and €364, respectively, among older people seeking Norwegian primary care due to a new episode of back pain. Patients

within the top 25th percentile accounted for 77% of all costs. Furthermore, patients classified as medium risk and high risk had a significantly higher degree of healthcare utilization compared with patients classified as low risk. Since this is the first study to estimate healthcare utilization and related cost among a sample of exclusively older people with back pain, further research is needed to complement these findings.

Twitter Ørjan Nesse Vigdal @orjanvigdal

Contributors RMK originated the idea. RMK, KS, DAW and MG designed the study. MG and KS contributed to the funding of the study. RMK, ZZK, ØNV and LK collected data for the study. RMK analysed the data. RMK, KS, DAW, ZZK, ØNV, MCS and MG contributed to the interpretation of data. RMK drafted the manuscript with all authors contributing in reading, commenting and approving the final manuscript. RMK is the guarantor.

Funding This study was supported by Oslo Metropolitan University, the Norwegian Fund for Post-Graduate Training in Physiotherapy (grant number 90749) and 'Et liv i bevegelse' (A life in movement)—Norwegian chiropractors' research foundation. Funding organisations had no part in the planning, performing, or reporting of the study.

Competing interests None declared.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Ethics approval The study was assessed by the Norwegian Regional Committee for Medical Research Ethics and was classified as a quality assessment study (ref number 2014/1634/REK vest). They specified that a quality assessment study does not require their explicit approval. The study was approved by the Norwegian Social Science Data Service (ref number 42149). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. Data not available.

ORCID iDs
Rikke Munk Killingmo http://orcid.org/0000-0003-1971-2449
Danielle van der Windt http://orcid.org/0000-0002-7248-6703
Ørjan Nesse Vigdal http://orcid.org/0000-0001-5075-7720
Margreth Grotle http://orcid.org/0000-0001-8243-1143

REFERENCES

1 Hartvigsen J, Hancock MJ, Kongsted A, *et al*. What low back pain is and why we need to pay attention. *Lancet* 2018;391:2356–67.
2 Buchbinder R, van Tulder M, Öberg B, *et al*. Low back pain: a call for action. *Lancet* 2018;391:2384–8.
3 Deyo RA, Mirza SK, Martin BI. Back pain prevalence and visit rates: estimates from U.S. national surveys, 2002. *Spine* 2006;31:2724–7.
4 Vos, T, GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the global burden of disease study 2019. *Lancet* 2020;396:1204–22.
5 Chen S, Chen M, Wu X, *et al*. Global, regional and national burden of low back pain 1990-2019: a systematic analysis of the global burden of disease study 2019. *J Orthop Translat* 2022;32:49–58.
6 van Tulder M, Koes B, Bombardier C. Low back pain. *Best Pract Res Clin Rheumatol* 2002;16:761–75.
7 Traeger AC, Buchbinder R, Elshaug AG, *et al*. Care for low back pain: can health systems deliver? *Bull World Health Organ* 2019;97:423–33.
8 Kinge JM, Knudsen AK, Skirbekk V, *et al*. Musculoskeletal disorders in Norway: prevalence of chronicity and use of primary and specialist health care services. *BMC Musculoskelet Disord* 2015;16:1–9.
9 Beyera GK, O'Brien J, Campbell S. Health-Care utilisation for low back pain: a systematic review and meta-analysis of population-based observational studies. *Rheumatol Int* 2019;39:1663–79.
10 Jöud A, Petersson IF, Englund M. Low back pain: epidemiology of consultations. *Arthritis Care Res* 2012;64:1084–8.
11 Licciardone JC. The epidemiology and medical management of low back pain during ambulatory medical care visits in the United States. *Osteopath Med Prim Care* 2008;2:11.
12 Werner ELIA. Kunnskap, praksis OG holdninger TIL rygglidelser HOS leger, fysioterapeuter OG kiropraktorer. *Tidsskr Nor Laegeforen* 2005;125:1794–7.
13 Kamper SJ, Logan G, Copsey B, *et al*. What is usual care for low back pain? A systematic review of health care provided to patients with low back pain in family practice and emergency departments. *Pain* 2020;161:694–702.
14 Foster NE, Anema JR, Cherkin D, *et al*. Prevention and treatment of low back pain: evidence, challenges, and promising directions. *Lancet* 2018;391:2368–83.
15 Corp N, Mansell G, Stynes S, *et al*. Evidence-Based treatment recommendations for neck and low back pain across Europe: a systematic review of guidelines. *Eur J Pain* 2021;25:275–95.
16 O'Connell NE, Cook CE, Wand BM, *et al*. Clinical guidelines for low back pain: a critical review of consensus and inconsistencies across three major guidelines. *Best Pract Res Clin Rheumatol* 2016;30:968–80.
17 Hill JC, Dunn KM, Lewis M, *et al*. A primary care back pain screening tool: identifying patient subgroups for initial treatment. *Arthritis Rheum* 2008;59:632–41.
18 NICE. *Low back pain and sciatica in over 16s: assessment and management*. London: National Institute for Clinical Excellence, 2016.
19 Hill JC, Whitehurst DGT, Lewis M, *et al*. Comparison of stratified primary care management for low back pain with current best practice (start back): a randomised controlled trial. *Lancet* 2011;378:1560–71.
20 Hall JA, Jowett S, Lewis M, *et al*. The start back stratified care model for nonspecific low back pain: a model-based evaluation of long-term cost-effectiveness. *Pain* 2021;162:702–10.
21 Maher C, Underwood M, Buchbinder R. Non-Specific low back pain. *Lancet* 2017;389:736–47.
22 Paeck T, Ferreira ML, Sun C, *et al*. Are older adults missing from low back pain clinical trials? A systematic review and meta-analysis. *Arthritis Care Res* 2014;66:1220–6.
23 Bressler HB, Keyes WJ, Rochon PA, *et al*. The prevalence of low back pain in the elderly. A systematic review of the literature. *Spine* 1999;24:1813–9.
24 Scheele J, Luijsterburg PAJ, Ferreira ML, *et al*. Back complaints in the elders (BACE); design of cohort studies in primary care: an international Consortium. *BMC Musculoskelet Disord* 2011;12:193.
25 Jarvik JG, Comstock BA, Bresnahan BW, *et al*. *Study protocol: the back pain outcomes using longitudinal data (BOLD) registry*. *BMC Musculoskelet Disord* 2012;13:64.
26 Hemingway H, Croft P, Perel P, *et al*. Prognosis research strategy (progress) 1: a framework for researching clinical outcomes. *BMJ* 2013;346:e5595.
27 Grotle M. *BACk Pain in Elders in Norway (BACE-N)*, 2020. Clinicaltrials.Gov. Available: https://clinicaltrials.gov/ct2/show/NCT04261309
28 Ø V. N, Storheim K, GM. With what accuracy can we predict persistent disability in sciatica patients with self-reported screening tools? *Oslo University College* 2014.
29 Bier JD, Ostelo RWJG, van Hooff ML, *et al*. Validity and reproducibility of the start back tool (Dutch version) in patients with low back pain in primary care settings. *Phys Ther* 2017;97:561–70.
30 Mehling WE, Avins AL, Acree MC, *et al*. Can a back pain screening tool help classify patients with acute pain into risk levels for chronic pain? *Eur J Pain* 2015;19:439–46.
31 Suri P, Delaney K, Rundell SD, *et al*. Predictive validity of the start back tool for risk of persistent disabling back pain in a U.S. primary care setting. *Arch Phys Med Rehabil* 2018;99:1533–9.
32 Rattay P, Butschalowsky H, Rommel A, *et al*. [Utilization of outpatient and inpatient health services in Germany: results of the German Health Interview and Examination Survey for Adults (DEGS1)]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2013;56:832–44.
33 Wenig CM, Schmidt CO, Kohlmann T, *et al*. Costs of back pain in Germany. *Eur J Pain* 2009;13:280–6.

34 Mutubuki EN, Luitjens MA, Maas ET, *et al*. Predictive factors of high societal costs among chronic low back pain patients. *Eur J Pain* 2020;24:325–37.
35 Chechulin Y, Nazerian A, Rais S, *et al*. Predicting patients with high risk of becoming high-cost healthcare users in Ontario (Canada). *Healthc Policy* 2014;9:68–79.
36 Hoebel J, Rattay P, Prütz F, *et al*. Socioeconomic status and use of outpatient medical care: the case of Germany. *PLoS One* 2016;11:e0155982.
37 Lim K-L, Jacobs P, Klarenbach S. A population-based analysis of healthcare utilization of persons with back disorders: results from the Canadian community health survey 2000-2001. *Spine* 2006;31:212–8.
38 Fritz JM, Kim J, Dorius J. Importance of the type of provider seen to begin health care for a new episode low back pain: associations with future utilization and costs. *J Eval Clin Pract* 2016;22:247–52.
39 Engel CC, von Korff M, Katon WJ. Back pain in primary care: predictors of high health-care costs. *Pain* 1996;65:197–204.
40 Ferreira ML, Machado G, Latimer J, *et al*. Factors defining care-seeking in low back pain--a meta-analysis of population based surveys. *Eur J Pain* 2010;14:747.e1–747.e7.
41 Becker A, Held H, Redaelli M, *et al*. Low back pain in primary care: costs of care and prediction of future health care utilization. *Spine* 2010;35:1714–20.
42 Lentz TA, Harman JS, Marlow NM, *et al*. Factors associated with persistently high-cost health care utilization for musculoskeletal pain. *PLoS One* 2019;14:e0225125.
43 Von Korff M, Jensen MP, Karoly P. Assessing global pain severity by self-report in clinical and health services research. *Spine* 2000;25:3140–51.
44 Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine* 1983;8:141–4.
45 Stewart WF, Yan X, Boscarino JA, *et al*. Patterns of health care utilization for low back pain. *J Pain Res* 2015;8:523–35.
46 Ritzwoller DP, Crounse L, Shetterly S, *et al*. The association of comorbidities, utilization and costs for patients identified with low back pain. *BMC Musculoskelet Disord* 2006;7:72.
47 Sangha O, Stucki G, Liang MH, *et al*. The self-administered comorbidity questionnaire: a new method to assess comorbidity for clinical and health services research. *Arthritis Rheum* 2003;49:156–63.
48 Ware JE, Sherbourne CD. The mos 36-item short-form health survey (SF-36). I. conceptual framework and item selection. *Med Care* 1992;30:473–83.
49 Keeley P, Creed F, Tomenson B, *et al*. Psychosocial predictors of health-related quality of life and health service utilisation in people with chronic low back pain. *Pain* 2008;135:142–50.
50 Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas* 1977;1:355–85.
51 Waddell G, Newton M, Henderson I, *et al*. A Fear-Avoidance beliefs questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain* 1993;52:157–68.
52 Enthoven WTM, Geuze J, Scheele J, *et al*. Prevalence and "Red Flags" Regarding Specified Causes of Back Pain in Older Adults Presenting in General Practice. *Phys Ther* 2016;96:305–12.
53 Verhagen AP, Downie A, Popal N, *et al*. Red flags presented in current low back pain guidelines: a review. *Eur Spine J* 2016;25:2788–802.
54 von Elm E, Altman DG, Egger M, *et al*. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med* 2007;4:e296.
55 Johnston KM, Lakzadeh P, Donato BMK, *et al*. Methods of sample size calculation in descriptive retrospective burden of illness studies. *BMC Med Res Methodol* 2019;19:9.
56 Zemedikun DT, Kigozi J, Wynne-Jones G, *et al*. Methodological considerations in the assessment of direct and indirect costs of back pain: a systematic scoping review. *PLoS One* 2021;16:e0251406.
57 Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. *Spine J* 2008;8:8–20.
58 Depont F, Hunsche E, Abouelfath A, *et al*. Medical and non-medical direct costs of chronic low back pain in patients consulting primary care physicians in France. *Fundam Clin Pharmacol* 2010;24:101–8.
59 Hong J, Reed C, Novick D, *et al*. Costs associated with treatment of chronic low back pain: an analysis of the UK general practice research database. *Spine* 2013;38:75–82.
60 Ekman M, Jönhagen S, Hunsche E, *et al*. Burden of illness of chronic low back pain in Sweden: a cross-sectional, retrospective study in primary care setting. *Spine* 2005;30:1777–85.
61 Olafsson G, Jonsson E, Fritzell P, *et al*. Cost of low back pain: results from a national register study in Sweden. *Eur Spine J* 2018;27:2875–81.
62 Lambeek LC, van Tulder MW, Swinkels ICS, *et al*. The trend in total cost of back pain in the Netherlands in the period 2002 to 2007. *Spine* 2011;36:1050–8.
63 Vlaeyen JWS, Maher CG, Wiech K, *et al*. Low back pain. *Nat Rev Dis Primers* 2018;4:52.
64 Wammes JJG, van der Wees PJ, Tanke MAC, *et al*. Systematic review of high-cost patients' characteristics and healthcare utilisation. *BMJ Open* 2018;8:e023113.
65 Werner EL, Ihlebæk C. Primary care doctors' management of low back pain patients--ten years after. *Tidsskr Nor Laegeforen* 2012;132:2388–90.
66 Patel ND, Broderick DF, Burns J, *et al*. ACR Appropriateness criteria low back pain. *J Am Coll Radiol* 2016;13:1069–78.
67 Henschke N, Maher CG, Refshauge KM, *et al*. Prognosis in patients with recent onset low back pain in Australian primary care: inception cohort study. *BMJ* 2008;337:a171.
68 Henschke N, Maher CG, Refshauge KM, *et al*. Prevalence of and screening for serious spinal pathology in patients presenting to primary care settings with acute low back pain. *Arthritis Rheum* 2009;60:3072–80.
69 Weiner DK, Kim Y-S, Bonino P, *et al*. Low back pain in older adults: are we utilizing healthcare resources wisely? *Pain Med* 2006;7:143–50.
70 Machado GC, Abdel-Shaheed C, Underwood M, *et al*. Non-Steroidal anti-inflammatory drugs (NSAIDs) for musculoskeletal pain. *BMJ* 2021;372:n104.
71 McCarberg BH. Nsaids in the older patient: balancing benefits and harms. *Pain Med* 2013;14 Suppl 1:S43–4.
72 Muller AE, Clausen T, Sjøgren P, *et al*. Prescribed opioid analgesic use developments in three Nordic countries, 2006-2017. *Scand J Pain* 2019;19:345–53.
73 Franklin M, Lomas J, Walker S, *et al*. An educational review about using cost data for the purpose of cost-effectiveness analysis. *Pharmacoeconomics* 2019;37:631–43.
74 Leurent B, Gomes M, Carpenter JR. Missing data in trial-based cost-effectiveness analysis: an incomplete journey. *Health Econ* 2018;27:1024–40.
75 Briggs A, Clark T, Wolstenholme J, *et al*. Missing presumed at random: cost-analysis of incomplete data. *Health Econ* 2003;12:377–92.
76 Petrou S, Murray L, Cooper P, *et al*. The accuracy of self-reported healthcare resource utilization in health economic studies. *Int J Technol Assess Health Care* 2002;18:705–10.
77 Icks A, Dittrich A, Brüne M, *et al*. Agreement found between self-reported and health insurance data on physician visits comparing different recall lengths. *J Clin Epidemiol* 2017;82:167–72.
78 Short ME, Goetzel RZ, Pei X, *et al*. How accurate are self-reports? analysis of self-reported health care utilization and absence when compared with administrative data. *J Occup Environ Med* 2009;51:786–96.
79 Hunger M, Schwarzkopf L, Heier M, *et al*. Official statistics and claims data records indicate non-response and recall bias within survey-based estimates of health care utilization in the older population. *BMC Health Serv Res* 2013;13:1.
80 Slagsvold B, Veenstra M, Daatland SO, *et al*. Life-course, ageing and generations in Norway:the NorLAG study. *Nor Epidemiol* 2012;22.
81 Vigdal Ørjan Nesse, Storheim K, Munk Killingmo R, *et al*. Characteristics of older adults with back pain associated with choice of first primary care provider: a cross-sectional analysis from the BACE-N cohort study. *BMJ Open* 2021;11:e053229.
82 Grotle M, Brox JI, Glomsrød B, *et al*. Prognostic factors in first-time care seekers due to acute low back pain. *Eur J Pain* 2007;11:290–8.
83 Nordstoga AL, Vasseljen O, Meisingset I, *et al*. Improvement in work ability, psychological distress and pain sites in relation to low back pain prognosis: a longitudinal observational study in primary care. *Spine* 2019;44:423–9.
84 Scheele J, Enthoven WTM, Bierma-Zeinstra SMA, *et al*. Characteristics of older patients with back pain in general practice: BACE cohort study. *Eur J Pain* 2014;18:279–87.
85 Nations U. World Population Ageing 2019. In: *Department of economic and social Affairs*. New York: Population Division, 2020.
86 Hofman A, Grobbee DE, de Jong PT, *et al*. Determinants of disease and disability in the elderly: the Rotterdam elderly study. *Eur J Epidemiol* 1991;7:403–22.

Paper II

Killingmo RM, Chiarotto A, van der Windt D, Storheim K,
Bierma-Zeinstra SMA, Småstuen MC, Zolic-Karlsson Z, Vigdal ØN, Koes BW, Grotle M

Modifiable prognostic factors of high costs related to healthcare utilization among older people seeking primary care due to back pain: an identification and replication study

BMC Health Serv Res. 2022 Jun 18;22(1):793
DOI: 10.1186/s12913-022-08180-2

BMC Health Services Research

RESEARCH

Open Access

Modifiable prognostic factors of high costs related to healthcare utilization among older people seeking primary care due to back pain: an identification and replication study

Rikke Munk Killingmo[1*], Alessandro Chiarotto[2], Danielle A. van der Windt[3], Kjersti Storheim[4,1], Sita M. A. Bierma-Zeinstra[2,5], Milada C. Småstuen[1], Zinajda Zolic-Karlsson[6], Ørjan N. Vigdal[1], Bart W. Koes[2,7] and Margreth Grotle[1,2]

Abstract

Background: Back pain is an extensive burden to our healthcare system, yet few studies have explored modifiable prognostic factors associated with high costs related to healthcare utilization, especially among older back pain patients. The aims of this study were to identify modifiable prognostic factors for high costs related to healthcare utilization among older people seeking primary care with a new episode of back pain; and to replicate the identified associations in a similar cohort, in a different country.

Methods: Data from two cohort studies within the BACE consortium were used, including 452 and 675 people aged ≥55 years seeking primary care with a new episode of back pain. High costs were defined as costs in the top 25th percentile. Healthcare utilization was self-reported, aggregated for one-year of follow-up and included: primary care consultations, medications, examinations, hospitalization, rehabilitation stay and operations. Costs were estimated based on unit costs collected from national pricelists. Nine potential modifiable prognostic factors were selected based on previous literature. Univariable and multivariable binary logistic regression models were used to identify and replicate associations (crude and adjusted for selected covariates) between each modifiable prognostic factor and high costs related to healthcare utilization.

Results: Four modifiable prognostic factors associated with high costs related to healthcare utilization were identified and replicated: a higher degree of pain severity, disability, depression, and a lower degree of physical health-related quality of life. Kinesiophobia and recovery expectations showed no prognostic value. There were inconsistent results across the two cohorts with regards to comorbidity, radiating pain below the knee and mental health-related quality of life.

Conclusion: The factors identified in this study may be future targets for intervention with the potential to reduce high costs related to healthcare utilization among older back pain patients.

Trial registration: ClinicalTrials.gov NCT04261309, 07 February 2020. Retrospectively registered.

Keywords: Back pain, Healthcare utilization, Costs, Prognostic factor research

*Correspondence: rikke@oslomet.no

[1] Department of Physiotherapy, Oslo Metropolitan University, Oslo, Norway
Full list of author information is available at the end of the article

Background

The burden of back pain has been growing along with an increasing and ageing population [1–5]. Back pain is the number one cause of disability globally [4] and an extensive burden to our healthcare systems [2, 6–8]. According to a recent systematic review, the prevalence rate of healthcare utilization for back pain ranges from 28 to 92% [9]. Back pain is one of the most prevalent complaints encountered in primary care [4, 5, 8, 10].

To improve use of scarce healthcare resources and reduce the burden on our healthcare systems, where possible and appropriate, researchers have highlighted the importance of monitoring and understanding healthcare utilization and related costs [3, 11]. It is well known that most of healthcare utilization and related costs stems from a relatively small group of back pain patients [12], and more importantly, that many of these patients receive unnecessary and ineffective treatment [3, 7]. This suggests that care for this high-cost subgroup requires quality improvement and cost reduction. An important next step towards this would be to identify modifiable prognostic factors associated with high costs related to healthcare utilization, and to replicate initial findings to evaluate the consistency of prognostic value across datasets and settings [13]. Information about such factors could inform development of effective strategies and/or interventions, or new applications of existing interventions.

Only a few prospective studies have explored modifiable prognostic factors associated with high costs related to healthcare utilization among patients with back pain [14–17], and no such study has been conducted among a sample of exclusively older people with back pain. Patients with high costs related to healthcare utilization are a diverse population [18, 19], and generalization of results cannot be done automatically from younger to older people with back pain [20]. With an ageing population, and the expected rise in older people requiring back care in the years to come [21], it is important to study modifiable prognostic factors of high costs related to healthcare utilization among older people with back pain.

Therefore, the aims of this study were 1) to identify modifiable prognostic factors for high costs related to healthcare utilization among older people seeking primary care with a new episode of back pain and 2) to replicate the identified associations in a similar cohort, in a different country.

Method

This study was designed and performed in accordance with the PROGnosis RESearch Strategy (PROGRESS) framework [22], with aims consistent with prognostic factor research: identification of prognostic factors, including external replication. In line with recommendations from the PROGRESS framework [13] a study protocol (ClinicalTrials.gov NCT04261309, 07 February 2020) including a statistical analysis plan has been published [23], and the REMARK reporting guidelines were followed [24].

Design and setting

This study was carried out in two steps. First, modifiable prognostic factors were identified in the Back Complaints in the Elderly - Norway study (BACE-N), a prospective observational cohort study within Norwegian primary care [25]. Next, a replication analysis was conducted in the Back Complaints in the Elders study (BACE-D), a prospective observational cohort study within Dutch primary care [26]. BACE-N has been classified as a quality assessment study by the Norwegian Regional Committee for medical Research Ethics (ref no. 2014/1634/REK vest) and approved by the Norwegian Social Science Data Service (ref no. 42149). Likewise, the BACE-D study protocol (NL24829.078.08) has been approved by the Medical Ethics Committee of the Erasmus Medical Center, the Netherlands. BACE-N and BACE-D are part of the international BACE consortium [26].

Participants and recruitment procedure

Eligible participants within BACE-N were people ≥55 years of age seeking primary care (physiotherapist, chiropractor, or General Practitioner (GP)) with a new episode of back pain (preceded by 6 months without visiting primary care for similar complaints). Eligible participants within BACE-D were people >55 years of age seeking primary care (GP) with a new episode of back pain (preceded by 6 months without visiting a GP for similar complaints). Patients were excluded from both studies if they had difficulties completing the questionnaires due to language barriers, or if they had difficulties completing the physical examination (e.g. are wheelchair bound). Patients within BACE-N were recruited by 110 physiotherapists, chiropractors and GPs in urban and rural parts of Norway between April 2015 and February 2020. Patients within BACE-D were recruited by 49 GPs in and around Rotterdam between March 2009 and September 2011. All included patients signed an informed consent form before study enrolment.

Data collection, outcome, modifiable prognostic factors, and covariates

At baseline all patients responded to a comprehensive questionnaire and went through a standardized physical examination. Follow-up questionnaires were sent at 3, 6, and 12 months after inclusion within BACE-N and at 3, 6, 9 and 12 months after inclusion within BACE-D. All

questionnaires were preferably completed electronically, but paper versions were available for patients not familiar with electronic data collection. Within this study, only data from questionnaires were used.

Outcome

The outcome of this study was costs related to healthcare utilization aggregated over one-year of follow up and dichotomized as high and low. Having high costs related to healthcare utilization was defined as patients with costs in the top 25th percentile [15, 16].

Healthcare utilization within BACE-N and BACE-D were self-reported and included: consultation to healthcare professionals (type and frequency), use of back medication (prescription and over-the-counter, type and frequency), number of diagnostic examinations (type and frequency), number of days of hospitalization and/or rehabilitation stay (within BACE-N) and back operations. Within BACE-N, consultations to healthcare professionals and use of back medication were reported with a 3-months recall period at each timepoint of follow-up. Number of diagnostic examinations and days of hospitalization and/or rehabilitation stay were reported with a 3-months recall period at 3- and 6-months follow-up, and a 6-months recall period at 12-months follow-up. Back operations were reported with a 12-months recall period at 12-months follow-up. Within BACE-D, all variables, except back operations, were reported with a 3-months recall period at each timepoint of follow-up. Back operations were reported with a 12-months recall period at 12-months follow-up. Total costs of healthcare utilization per patient were estimated by multiplying frequency of use by unit costs collected from national pricelists (see Table 1).

Modifiable prognostic factors

Potential modifiable prognostic factors were factors expected to have the potential to be modified or improved by appropriate care or treatment, and therefore classified as modifiable. Potential modifiable prognostic factors of high costs related to healthcare utilization were based on previous scientific literature on (primarily) middle-aged back pain patients as well as patients with musculoskeletal disorders, and included the following self-reported variables measured at baseline:

- Pain severity [14–16, 28–30] measured by a Numeric Rating Scale (NRS) (range 0-10, higher score indicating higher pain severity) [31].
- Disability [14–16, 28, 29, 32] measured by the Roland-Morris Disability questionnaire (RMDQ) (range 0-24, higher score indicating higher degree of back-related disability) [33].
- Health-related quality of life [14, 30] measured by the Short-Form Health Survey 36-item (SF36) physical and mental summary score (range 0-100, higher score indicating better health-related quality of life) [34].
- Emotional well-being [15, 16, 19, 32, 35] measured by the Center for Epidemiological Studies-Depression questionnaire (CES-D) (range 0-60, higher score indicating more signs of depression) [36].
- Kinesiophobia [15, 35] measured by the Fear Avoidance Beliefs Questionnaire - Physical Activity subscale (FABQ-PA) (range 0-24, higher score indicating higher levels of kinesiophobia) [37].
- Comorbidity [17, 30] measured by the Self-Administered Comorbidity Questionnaire (SCQ) (range 0-15, thirteen pre-defined comorbidities and two optional comorbidities. Item no. 12 (back pain) was replaced with a third optional comorbidity) [38].
- Radiating pain below the knee [15] measured by the question "did your back pain radiate to your legs last week? If yes, how far down did the pain radiate?" and categorized into yes/no.
- Expectations of recovery from back pain within the next 3 months measured with a five-point scale and categorized into "recovered", "much better" or "no change or worse".

Covariates

Prognostic factor research may vary depending on context (time, place, healthcare setting) and characteristics of the study population. We therefore adjusted for potential covariates when evaluating the modifiable prognostic factors. Potential covariates were based on previous scientific literature on (primarily) middle-aged back pain patients as well as patients with musculoskeletal disorders, and included the following self-reported variables measured at baseline:

- Sex [14, 28, 39, 40] (female/male).
- Age [14, 28, 39, 40] (years).
- Education level [32, 41] measured as the highest education completed and categorised into low (elementary and high school level) or high (university level).
- Employment status measured by the question "do you have a paying job?" and categorized into yes/no.
- Pain duration [16] measured by the question "how many days have you had your current back pain?" and categorized into < 6 weeks, 6 weeks to 3 months or > 3 months.
- Pain history [29] measured by the question "have you had back pain before?" and categorized into yes/no.

Table 1 Cost categories, units, unit price, all numbers in Euros (€) for 2020

Cost categories	Unit	Norwegian unit price (€)	Dutch unit price (€)	Reference (source)
Primary care				
General practitioner	Per visit	43.1	36.0	The Norwegian Medical Association, estimated average iMTA costing tool [27]
Occupational physician	Per visit	–	36.0	iMTA costing tool [27]
Physiotherapist	Per visit	47.2	36.0	The Norwegian Physiotherapy Association, estimated average iMTA costing tool [27]
Chiropractor	Per visit	55.0	36.0	Private price lists, estimated average. iMTA costing tool [27]
Manuel therapist	Per visit	74.2	36.0	The Norwegian Physiotherapy Association, estimated average iMTA costing tool [27]
Naprapath	Per visit	64.0	–	Private price lists, estimated average
Osteopath	Per visit	65.0	–	Private price lists, estimated average
Psychologist	Per visit	110.0	102.0	The Norwegian Psychological Association, estimated average iMTA costing tool [27]
Other therapists	Per visit	75.0	–	Private price lists, estimated average
Back medication				
Paracetamol	Per daily defined dose	0.5	0.9	NoMA price list, estimated average. Medicijnkosten.nl, estimated average incl. Pharmacy delivering costs [27]
NSAID	Per daily defined dose	1.2	0.4	NoMA price list, estimated average. Medicijnkosten.nl, estimated average incl. Pharmacy delivering costs [27]
Muscle relaxant	Per daily defined dose	0.7	0.5	NoMA price list, estimated average. Medicijnkosten.nl, estimated average incl. Pharmacy delivering costs [27]
Sleep medication	Per daily defined dose	0.2	–	NoMA price list, estimated average
Cortisone	Per daily defined dose	0.4	–	NoMA price list, estimated average
Opioid	Per daily defined dose	0.9	0.5	NoMA price list, estimated average. Medicijnkosten.nl, estimated average incl. Pharmacy delivering costs [27]
Antidepressant	Per daily defined dose	–	0.3	Medicijnkosten.nl, estimated average incl. Pharmacy delivering costs [27]
Anticonvulsant	Per daily defined dose	–	0.7	Medicijnkosten.nl, estimated average incl. Pharmacy delivering costs [27]
Examinations				
Blood sample	Per examination	20.4	4.4	The Norwegian Medical Association, estimated average iMTA costing tool [27]
X-ray	Per examination	119.0	45.9	Unilabs price list, estimated average The National Health Authority
MRI	Per examination	269.0	233.0	Unilabs price list, estimated average iMTA costing tool [27]
CT	Per examination	189.0	151.0	Unilabs price list, estimated average iMTA costing tool [27]
Secondary care				
Medical specialist	Per visit	–	125.0	iMTA costing tool [27]
Back operation	Per operation	5220.0	5254.0	DRG2150. Different academic and non-academic hospitals pricelists, estimated average
Hospitalization (non-operation)	Per day	1880.0	–	The Norwegian Directorate of Health, SAMDATA
Rehabilitation stay	Per day	315.0	–	UniCare pricelist, estimated average

iMTA indicates institute for Medical Technology Assessment, *NSAID* Non-steriodal anto-anflammatory drug, *NoMA* Norwegian Medicines Agency. Cells marked with a dash (–) indicate that the unit price was not estimated

- First healthcare provider [42] (physiotherapist, chiropractor, or GP).
- Total costs related to healthcare utilization during a period of 6 (BACE-N) or 12 (BACE-D) weeks prior to inclusion. Healthcare utilization prior to inclusion was self-reported and included: primary care consultations, use of back medication and number of diagnostic examinations. Total cost of healthcare utiliza-

tion was estimated by multiplying frequency of use by unit costs collected from national pricelists (see Table 1).

Analyses

All analyses are outlined in the statistical analysis plan published a priori [25] and preformed using the IBM SPSS version 26 (IBM Corporation, Armonk, NY, USA). We considered our study as explanatory. Thus, no correction for multiple testing was performed and p-values < 0.05 were considered statistically significant. All statistical tests were two-sided.

Study flow

The flow of patients through the studies were reported with a flow chart according to the REMARK guidelines [24]. Reasons for dropout were provided where known. Dropouts at 12-months follow-up were removed from the analyses. Differences in baseline characteristics between responders and non-responders at 12-months follow-up were evaluated.

Missing data

Whitin BACE-N, missing value pattern was visually explored, and missingness at random was assumed. Also, we found evidence against the hypothesis that values were not missing completely at random (Little's test, $p > 0.05$). Missing baseline data was handled by multiple imputation. Five multiple imputation datasets with 10 iterations were created using regression estimation. We did not impute missing outcome values, as the imputation model had poor predictive performance and caused a clear trend of values being overestimated. Instead, missing values on variables used to estimate the outcome score were filled in with; 1) each patient's individual average of observed values for the variables: consultations to healthcare professionals and medication use, 2) a value of zero costs for the variables: diagnostic examinations, hospitalization, rehabilitation stay and back operations. Within BACE-D, missing value pattern was visually explored, and missingness at random was assumed. Missing values on variables used to estimate the outcome score were filled in with; 1) each patient's individual average of observed values for the variables: consultations to healthcare professionals and medication use, 2) a value of zero costs for the variables: diagnostic examinations and back operations.

Healthcare utilization and cost estimation

Type and frequency of use of different healthcare resources were calculated for each of the follow-up periods. Costs of healthcare utilization per patient were estimated by multiplying frequency of use by unit prices collected from national pricelists (see Table 1). Costs related to back medication were estimated based on medication type and frequency of use (data on dosage were not available). All costs were presented in Euros (€) for 2020 and estimated for the entire follow-up period with both mean and median values with 95% CI, using bias-corrected and accelerated (BCa) bootstrapping. The BCa was conducted with a bootstrap sample size of 1000. Cost data are commonly skewed thus both mean and median values were presented to inform interpretation. Norwegian prices were recalculated to Euros using the exchange rate from the National Bank of Norway from February 2020 (1€ = NOK 10).

Identification analysis

Univariable and multivariable binary logistic regression models were used to investigate associations (crude and adjusted for selected covariates) between each predefined modifiable prognostic factor and costs related to healthcare utilization (within BACE-N). The cost score was entered into the model as a dependent dichotomous variable (high costs defined as patients with cost in the top 25th percentile, yes/no). Linearity of continuous independent variables were examined using Box-Tidwell transformations [43]. Independent variables that demonstrated a non-linear relationship with the dependent variable where categorized. The results were presented as crude and adjusted odds ratios (OR) with 95% CI.

Replication analysis

Univariable and multivariable binary logistic regression models were used, as described above, to replicate findings from the identification analysis within BACE-D. The results were presented as crude and adjusted OR with 95% CI. The decision on whether findings were replicated were based on the direction and magnitude of the association, and the size of the CI for each of the predefined modifiable prognostic factors [44].

Sensitivity analysis

To assess credibility of the identification analysis and possible bias introduced by the imputation procedure, the univariable and multivariable logistic regression analyses were performed on complete case data (within BACE-N).

Sample size

This study contains secondary analyses embedded in the BACE-N and BACE-D. Details of the sample size calculation related to the original aims of the cohorts are provided in the BACE-N and BACE-D protocols [25, 26]. To determine statistical power for this particular study, we used number of events per variable (EPV) [45–49] and

the rule-of-thumb of "10 events per 1 analysed variable" [50–53]. With a sample size of 450 participants within BACE-N, we anticipated 112 participants to be in the top 25th percentile of costs and categorized as having high costs (yes/no) (events). An EPV of 10 would allow a maximum of 11 prognostic variables to be included in the final multivariable prediction model. With a sample size of 675 participants in BACE-D, we anticipated 168 participants to be in the top 25th percentile of costs and defined as having high costs (yes/no) (events). An EPV of 10 would allow a maximum of 16 prognostic variables to be included in the final multiple prediction model.

Results

A total of 452 (BACE-N) and 675 (BACE-D) patients were included in the identification and the replication sample, respectively. Table 2 shows patient characteristics and clinical status at baseline, along with the proportion of missing data per variable. Flow of patients through the studies are shown in Fig. 1. Fourteen patients (3%) in BACE-N and 22 patients (3%) in BACE-D were drop-outs at 12-months follow-up. We removed these cases from the analyses. Within BACE-N, there was a larger proportion of females (55 vs. 42%) among the responders as compared to non-responders. Within BACE-D, there was a larger proportion of people not in paid work (26 vs. 38%) and people with short pain duration < 6 weeks (56 vs. 39%) among the responders as compared to non-responders. Otherwise, there were no differences between responders and non-responders in the two cohorts. The BACE-N and BACE-D samples were also largely comparable, although there were some differences that might have impacted healthcare utilization. BACE-N had a larger proportion of people with high education level (44 vs. 17%), people in paid work (47 vs. 27%), and people with short pain duration < 6 weeks (67 vs. 54%).

Within BACE-N, missing data ranged from 0.0 to 16.8% for included baseline variables and 18.4 to 26.0% for included follow-up variables. Total missingness was 4.9 and 23.3% for all baseline and follow-up values, respectively. Variables on medication use at 12-months follow-up had most missing values. Within BACE-D, missing data ranged from 0.0 to 11.8% for included baseline variables and 7.3 to 18.1% for included follow-up variables. Total missingness was 2.2 and 9.6% for all baseline and follow-up values, respectively. Variables on examination at 12-months follow-up had most missing values.

Healthcare utilization and cost estimation

Table A1 and A2 in the Additional file 1 and Additional file 2 shows healthcare utilization throughout one-year of follow-up for the BACE-N and BACE-D sample, respectively. Costs related to healthcare utilization aggregated for the one-year of follow-up are shown in Table 3. Within BACE-N, 87% of all patients used healthcare during the one-year of follow-up, and a total of 110 patients (25%) were defined as having high costs (≥ € 789). Within BACE-D, 78% of all patients used healthcare during the one-year of follow-up, and a total of 163 patients (25%) were defined as having high costs (≥ € 664).

Identification analysis

All continuous independent variables, aside from the two SF-36 variables, demonstrated a linear relationship with the dependent variable. Table 4 shows crude and adjusted OR with 95% CI for the association between each of the modifiable prognostic factor and being in the high costs group. All analyses showed a statistical significant crude association between the factors and the outcome. After adjustment for covariates, only the following factors remained significantly associated with the outcome: pain severity, disability, depression, comorbidity, radiating pain below the knee, and physical and mental health-related quality-of-life.

The sensitivity analysis (Table A3 in the Additional file 3) showed no substantial change in point estimates when comparing complete case analysis to the main analysis. There were some minor changes in *p*-values for the two SF-36 variables: In the complete case analysis of crude associations, the SF-36 mental second percentile group were not significantly associated with the outcome, and in the complete case analysis of adjusted associations, the SF-36 physical and mental second percentile groups were not significantly associated with the outcome.

Replication analysis

Table 4 also shows results of the replication analysis. Except for the SF-36 mental second percentile group, findings were replicated with respect to the direction of the association between each of the factors and the outcome. Though, the magnitude of the association varied > 20% for the following factors: comorbidity, radiating pain below the knee, and physical and mental health-related quality-of-life.

In both the identification and replication analysis, after adjustment for selected covariates and with the "low cost group" as the reference, factors associated with increased odds of being in the high costs group were a higher degree of pain severity, disability and depression, and a lower degree of physical health-related quality of life. No association was found between being in the high costs group and the degree of kinesiophobia or expectations of recovery.

Table 2 Patient characteristics and clinical status at baseline in the identification and replication sample*

	BACE-N		BACE-D	
	All participants (*n* = 452)	**Missing, n (%)**	**All participants (*n* = 675)**	**Missing, n (%)**
Female	235 (52)	0 (0)	401 (59)	0 (0)
Age in years	66 (59-72)	0 (0)	65 (60-71)	0 (0)
Education level high	188 (44)	20 (4)	114 (17)	7 (1)
Ethnicity Norwegian (BACE-N) or Dutch (BACE-D)	430 (95)	0 (0)	637 (96)	10 (1)
Employment status currently paid work	211 (47)	5 (1)	177 (27)	23 (3)
First healthcare provider				
General practitioner	127 (28)	0 (0)	675 (100)	0 (0)
Physiotherapist	130 (29)	0 (0)	0 (0)	0 (0)
Chiropractor	195 (43)	0 (0)	0 (0)	0 (0)
Pain location				
Thoracic	56 (13)	11 (2)	154 (26)	71 (11)
Lumbar/Sacral	406 (92)	11 (2)	561 (93)	71 (11)
Radiating pain below the knee	141 (31)	0 (0)	205 (31)	7 (1)
Pain severity average last week (NRS, 0-10)	5 (4-7)	31 (7)	5 (3-7)	11 (2)
Pain duration				
< 6 weeks	252 (67)	76 (17)	323 (54)	80 (12)
6 weeks to 3 months	49 (13)	76 (17)	116 (20)	80 (12)
> 3 months	75 (20)	76 (17)	156 (26)	80 (12)
Previous episodes of back pain	400 (95)	29 (6)	579 (86)	9 (1)
Disability (RMDQ, 0-24)	9 (4-13)	45 (10)	10 (5-14)	55 (8)
Comorbidity (SCQ, 0-15)	1 (1-2)	18 (4)	2 (1-3)	6 (1)
Health-related QOL (SF36, 0-100)				
Physical component	42 (36-47)	41 (9)	43 (37-50)	7 (1)
Mental component	55 (48-60)	41 (9)	52 (43-57)	7 (1)
Emotional well-being (CES-D, 0-60)	8 (3-13)	57 (13)	9 (4-14)	57 (8)
Kinesiophobia (FABQ-PA, 0-24)	9 (5-13)	18 (4)	14 (10-17)	20 (3)
Expectations of recovery within 3 months				
Fully recovered	111 (26)	19 (4)	113 (17)	17 (2)
Much better	217 (50)	19 (4)	178 (27)	17 (2)
No change or worse	105 (24)	19 (4)	367 (56)	17 (2)
Healthcare utilization prior to inclusion				
Primary care consultation last 6 (BACE-N) or 12 (BACE-D) weeks				
General practitioner	78 (18)	21 (5)	609 (91)	8 (1)
Occupational physician	–	–	13 (2)	8 (1)
Physiotherapist, Chiropractor or Manual therapist	295 (68)	21 (5)	299 (45)	8 (1)
Psychologist	2 (0.5)	21 (5)	5 (1)	8 (1)
Other therapists	21 (5)	21 (5)	–	–
Use of back medication	165 (40)	38 (8)	484 (73)	8 (1)
Diagnostic examination last 6 (BACE-N) or 3 (BACE-D) months				
Blood sample	12 (3)	24 (5)	92 (14)	10 (2)
X-ray	23 (5)	24 (5)	155 (23)	10 (2)
MRI/CT scan	49 (11)	24 (5)	30 (5)	10 (2)
Previous hospitalization	48 (11)	21 (5)	–	–
Previous rehabilitation stay	17 (4)	25 (6)	–	–
Medical specialist consultation	–	–	46 (7)	8 (1)

CES-D indicates The Center for Epidemiologic Studies Depression Scale, *FABQ-PA* The Fear Avoidance Beliefs Questionnaire, physical activity subscale, *NRS* Numeric Rating Scale, *RMDQ* The Roland Morris Disability Questionnaire, *SCQ* The Self-Administered Comorbidity Questionnaire, *SF-36* The Short-Form Health Survey 36-item.
*The presented characteristics are based on complete case data. All values are presented by number (valid percentage of total) or median (IQR). Cells marked with a dash (–) indicate that the variable was not measured

Table 3 Costs (€) due to healthcare utilization from 0 to 12 month in the identification and replication sample*

	BACE-N ($n = 438$)			BACE-D ($n = 653$)		
	Mean (95% CI)**	**Median (95% CI**)**	**Patients with zero cost, n (%)**	**Mean (95% CI**)**	**Median (95% CI**)**	**Patients with zero cost, n (%)**
Primary care	458 (404-516)	242 (192-330)	83 (21)	289 (255-329)	72 (72-72)	250 (40)
Medication	52 (43-61)	3 (1-7)	176 (44)	62 (54-70)	7 (0-17)	291 (46)
Examination	65 (50-81)	0 (0-0)	308 (77)	73 (63-87)	0 (0-4)	342 (54)
Secondary care	243 (116-388)	0 (0-0)	390 (97)	158 (110-213)	0 (0-0)	503 (80)
Total	825 (682-976)	364 (307-440)	52 (13)	582 (506-666)	233 (190-276)	136 (22)

*Costs due to healthcare utilization for the entire follow-up period is calculated on basis of the three (BACE-N) and four (BACE-D) follow-up periods. Costs in the two cohorts are not direct comparable. The BACE-N lack data on primary care consultations and medication use between 6 and 9 months. Thus, total costs within the BACE-N are expected to be slightly underestimated. **Bias-corrected and accelerated bootstrapping (1000 simulations)

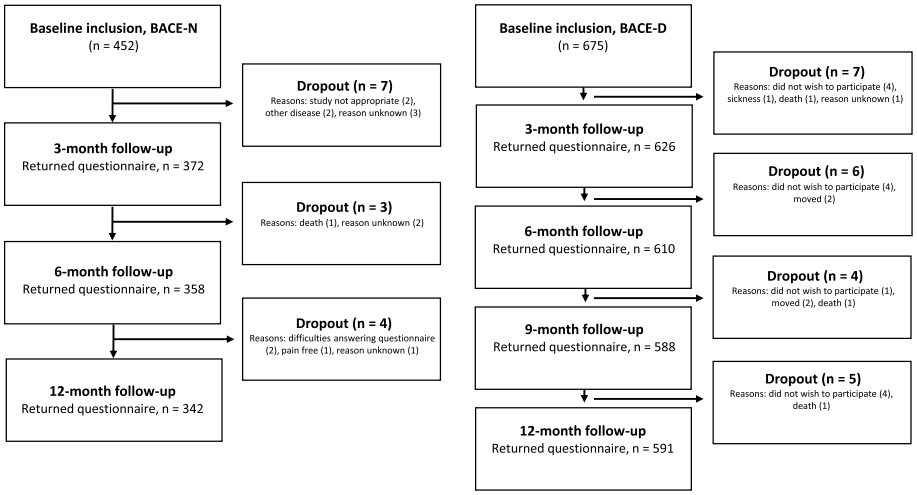

Fig. 1 Participant flow chart BACE-N and BACE-D

Discussion

The present study identified and replicated associations between modifiable prognostic factors and high costs related to healthcare utilization among older people seeking primary care with a new episode of back pain. Four modifiable prognostic factors associated with high costs related to healthcare utilization were identified and replicated in a similar cohort, in a different country, reflecting slightly different sociodemographic characteristics and healthcare setting: pain severity, disability, depression and physical health-related quality of life. Kinesiophobia and expectations of recovery showed no prognostic value. There were inconsistent results across the two cohorts with regards to comorbidity, radiating pain below the knee and mental health-related quality of life.

To the best of our knowledge, no similar study has been conducted among a sample of exclusively older people with back pain. Thus, direct comparability of this study with other studies is limited. Nevertheless, our findings are generally in accordance with previous research on (primarily) middle-aged back pain patients [14–16, 28, 29, 32, 35], as well as patients with musculoskeletal disorders [30, 54, 55]. For example, pain severity, disability and depression have been shown to be significantly associated with high costs related to healthcare utilization in studies on patients with back pain [14–16, 28, 29, 32, 35] and musculoskeletal disorders [30, 54, 55]. Physical health-related quality of life has also previously been reported to be a prognostic factor of high societal costs among back pain patients [14], and high costs related to healthcare utilization among patients with musculoskeletal disorders [30]. Our findings regarding kinesiophobia and radiating pain below the knee are also in line with a previous study [15], which showed that these factors were

Table 4 Binary logistic regression analyses; individual associations between modifiable prognostic factors and high costs

	BACE-N ($n = 438$)		BACE-D ($n = 653$)	
	Crude OR (95% CI)	Adjusted OR* (95% CI)	Crude OR (95% CI)	Adjusted OR* (95% CI)
Pain severity (NRS, 0-10)	1169 (1.059-1.291)	1.147 (1.031-1.277)	1.295 (1.198-1.400)	1.324 (1.203-1.457)
Disability (RMDQ, 0-24)	1146 (1.096-1.198)	1.140 (1.087-1.195)	1.143 (1.101-1.186)	1.143 (1.092-1.196)
Emotional well-being (CES-D, 0-60)	1.050 (1.024-1.076)	1.040 (1.013-1.068)	1.041 (1.017-1.065)	1.038 (1.010-1.066)
Kinesiophobia (FABQ-PA, 0-24)	1.050 (1.012-1.090)	1.030 (0.990-1.071)	1.037 (1.005-1.071)	1.034 (0.997-1.073)
Comorbidity (SCQ, 0-15)	1.611 (1.363-1.905)	1.614 (1.339-1.945)	1.179 (1.051-1.323)	1.091 (0.948-1.254)
Radiating pain below knee (ref: no)	2.604 (1.662-4.080)	2.254 (1.389-3.660)	2.124 (1.468-3.073)	1.507 (0.969-2.345)
Health-related QOL physical (SF36, 0-100) (ref. 4. percentile)				
3. percentile	2.731 (1.237-6.029)	2.167 (0.956-4.909)	2.876 (1.649-5.016)	2.328 (1.222-4.437)
2. percentile	3.836 (1.774-8.296)	2.778 (1.250-6.173)	2.406 (1.440-4.021)	2.198 (1.221-3.958)
1. percentile	7.185 (3.377-15.290)	4.913 (2.235-10.803)	4.326 (2.562-7.303)	3.937 (2.082-7.445)
Health-related QOL mental (SF36, 0-100) (ref. 4. percentile)				
3. percentile	0.917 (0.452-1.859)	1.095 (0.523-2.292)	0.861 (0.473-1.556)	0.813 (0.410-1.613)
2. percentile	2.092 (1.104-3.961)	2.162 (1.102-4.240)	0.503 (0.272-0.933)	0.382 (0.187-0.784)
1. percentile	2.717 (1.444-5.113)	2.583 (1.317-5.068)	1.375 (0.779-2.367)	1.173 (0.629-2.185)
Expectations of recovery within 3 months (ref. recovered)				
Much better	2.321 (1.300-4.144)	1.622 (0.878-2.997)	1.552 (0.849-2.840)	1.129 (0.583-2.186)
No change or worse	1.547 (0.784-3.053)	1.004 (0.483-2.087)	1.887 (1.093-3.257)	1.275 (0.680-2.391)

CES-D indicates The Center for Epidemiologic Studies Depression Scale, *CI* Confidence interval, *FABQ-PA* The Fear Avoidance Beliefs Questionnaire, physical activity subscale, *NRS* Numeric Rating Scale, *OR* Odds ratio, *RMDQ* The Roland Morris Disability Questionnaire, *SCQ* The Self-Administered Comorbidity Questionnaire, *SF-36* The Short-Form Health Survey 36-item. *Adjusted by sex, age, education level, employment status, pain duration, pain history, first healthcare provider and costs related to healthcare utilization prior to inclusion

of minor importance when predicting future costs related to healthcare utilization among back pain patients. Our finding regarding mental health-related quality of life is, however, contrary to a study on patients with musculoskeletal disorders [30], which found that this factor was associated with persistent high costs related to healthcare utilization. Our finding regarding comorbidity is also contrary to previous research. In a recent systematic review, comorbidity was pointed out as a consistent prognostic factor of high costs related to healthcare utilization in general [18], and similar conclusions have been drawn in single studies among patients with back pain [17] and musculoskeletal disorders [55]. This discrepancy might be explained by the fact that we included costs related to back pain specific healthcare utilization, whereas other studies have included healthcare costs related to all musculoskeletal disorders [17, 55] and healthcare costs in general [18]. To the best of our knowledge, the prognostic value of recovery expectations for high costs related to healthcare utilization has not been reported previously.

The main limitation of this study is missing data on variables used to estimate the outcome score, thus we had to manually replace missing values. It is well-known that healthcare utilization is prone to missing data [56–58]. Also, that missing data should be replaced in order to make use of all reported data [56, 57]. Unfortunately, due to poor predictive performance, multiple imputation could not be used on follow-up data in this study. We therefore chose a frequently used, though not optimal, method for replacing missing values [58] and have tried to be transparent in our reporting. A second potential limitation is that we used self-reported data on healthcare utilization. Self-reports tend to underestimate the true value of healthcare utilization due to potential recall bias [59–62]. Nevertheless, we consider the impact of recall bias to be of only minor importance in this study as the outcome variable was dichotomized into high or low costs. In future studies, the limitations of missing data and recall bias could to some extent be overcome by including registry data on healthcare utilization. A third potential limitation is that costs related to hospitalization and rehabilitation stays were not measured in BACE-D. Thus, the risk of misclassification bias related to whether patients were classified as having high or low costs might be present in BACE-D. However, if costs related to hospitalization and rehabilitation stays were removed from the cost calculations in BACE-N, only 4 patients (< 1%) switched cost group. A fourth potential limitation is that we could not adjust for possibly important covariates of healthcare utilization, such as the patient's disposition to access and pay for healthcare services, and health insurance status. According to the Behavioral Model of Health Services Use from Andersen [63], healthcare utilization is a function of people's predisposition to use services,

factors which enable or impede use and need for care. Certainly, including these enabling factors is recommended. However, it is likely to assume that these factors are of less importance in countries such as Norway and the Netherlands, where health services are largely available and covered by the public sector. A fifth limitation is the lack of data on eligible participants that declined to participate or for other reasons were not invited. Due to limited resources and practical reasons related to recruitment from a broad network of clinicians, it was not possible to record information on all eligible participants during the BACE-N and BACE-D data collection period. Thus, the risk of selection bias is present. To compensate for this limitation and assess representativeness of the BACE-N sample, key sociodemographic variables have been compared with a large population study on older people; The Norwegian study on life course, ageing and generation (NORLAG) [64, 65]. A subsample of the NORLAG (NORLAG MSK) was used, which is expected to be a representative sample of people aged ≥55 years with musculoskeletal complaints. Characteristics of the two samples were largely comparable, though BACE-N had more men, and more with higher education level [65]. Previous studies have shown that women [14, 28, 29] are more likely to seek care for their back pain as are people with lower education level [28, 32, 41]. Hence, it is likely to assume that the amount of healthcare utilization presented in BACE-N is somewhat underestimated. Furthermore, the BACE-N sample is largely comparable to younger Norwegian back pain cohorts [66, 67] and to the BACE-D sample [26].

The main strength of the present study is that it was conducted in line with the PROGRESS framework including an identification and replication phase [13], pre-planned with a published statistical analysis plan, and reported in line with the REMARK guidelines [24]. Also, that it estimates the prognostic value of modifiable prognostic factors over and beyond a core set of non-modifiable covariates. Prognostic factor studies are an essential step towards quality improvement of clinical practice [13]. Results from such studies have the potential to inform development of effective strategies and/or interventions. Identifying modifiable prognostic factors of high costs related to healthcare utilization among older people is an important step towards addressing the global burden of back pain and decrease waste of valuable healthcare resources [3, 7, 68].

Conclusion

In conclusion, this study identified and replicated four modifiable prognostic factors associated with high costs related to healthcare utilization among older people seeking primary care with a new episode of back pain: pain severity, disability, depression, and physical health-related quality of life. This study contributes to the on-going research into clinical pathways and has the potential to identify future target areas for intervention with the potential to reduce high costs related to healthcare utilization among older back pain patients. Due to differences in healthcare systems between countries, readers are advised to exercise caution with generalizability of the results to other healthcare systems.

Abbreviations

BCa: Bias-corrected and accelerated; BACE-D: Back Complaints in the Elders study; BACE-N: Back Complaints in the Elderly - Norway study; CES-D: Center for Epidemiological Studies-Depression questionnaire; CI: Confidence Interval; EPV: Events per variable; FABQ-PA: Fear Avoidance Beliefs Questionnaire - Physical Activity subscale; GP: General Practitioner; NOK: Norwegian krone; NORLAG: Norwegian study on life course, ageing and generation; NRS: Numeric Rating Scale; OR: Odds ratio; PROGRESS: PROGnosis RESearch Strategy; RMDQ: Roland-Morris Disability Questionnaire; REMARK: REporting recommendations for tumor MARKer prognostic studies; SCQ: Self-Administered Comorbidity Questionnaire; SF36: Short-Form Health Survey 36-item.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12913-022-08180-2.

Additional file 1.

Additional file 2.

Additional file 3.

Acknowledgements

We would like to thank all study participants, participating physiotherapists, chiropractors and GPs and their assistants for their contribution to the study. We thank the primary care providers assisting with the clinical examinations and providing baseline questionnaires to study participants. We thank the BACE-N scientific board, and the BACE Consortium.

Authors' contributions

RMK originated the idea. RMK, DAW, KS, BWK and MG designed the study. MG and KS contributed to the funding of the study. RMK, ZZK and ØNV collected data for the study. RMK and AC analysed the data. RMK, AC, DAW, KS, SMABZ, MCS, ZZK, BWK and MG contributed to the interpretation of data. RMK drafted the manuscript with all authors contributing in reading, commenting and approving the final manuscript.

Funding

This study was supported by the Oslo Metropolitan University, the Norwegian Fund for Post-Graduate Training in Physiotherapy (grant number 90749) and "Et liv i bevegelse" (A life in movement) - Norwegian chiropractors' research foundation. Funding organisations had no role in the design of the study or collection, analysis, and interpretation of the data or in writing the manuscript.

Availability of data and materials

Data supporting findings of this study are not public available as participants have consented for their data to be available only to the researcher of this study. However, data are available from the corresponding author upon reasonable request and with permission of the Oslo Metropolitan University and the Erasmus Medical Center (contact through the corresponding author).

Declarations

Ethics approval and consent to participate

The BACE-N study was classified as a quality assessment study by the Norwegian Regional Committee for Medical Research Ethics (ref no. 2014/1634/REK vest) and approved by the Norwegian Social Science Data Service (ref no. 42149). The BACE-D study was approved by the Medical Ethics Committee of the Erasmus Medical Center, the Netherlands (NL24829.078.08). This study was performed in accordance with the PROGRESS framework and the Declaration of Helsinki. All participants provided written informed consent to participate in the studies.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Physiotherapy, Oslo Metropolitan University, Oslo, Norway. [2]Department of General Practice, Erasmus MC, University Medical Centre, Rotterdam, The Netherlands. [3]School of Medicine, Keele University, Keele, UK. [4]Research and Communication Unit for Musculoskeletal Health (FORMI), Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway. [5]Department of Orthopedics, Erasmus MC, University Medical Centre, Rotterdam, The Netherlands. [6]HTA and Reimbursement at Norwegian Medicines Agency, Oslo, Norway. [7]Center for Muscle and Joint Health, University of Southern Denmark, Odense, Denmark.

Received: 17 December 2021 Accepted: 13 June 2022
Published online: 18 June 2022

References

1. Vos T, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2163–96.
2. Hartvigsen J, et al. What low back pain is and why we need to pay attention. Lancet. 2018;391(10137):2356–67.
3. Buchbinder R, et al. Low back pain: a call for action. Lancet. 2018;391(10137):2384–8.
4. Diseasse GBD, Injury I, Prevalence C. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016;388(10053):1545–602.
5. Deyo RA, Mirza SK, Martin BI. Back pain prevalence and visit rates: estimates from U.S. national surveys. Spine (Phila Pa 1976), 2006. 2002;31(23):2724–7.
6. van Tulder M, Koes B, Bombardier C. Low back pain. Best Pract Res Clin Rheumatol. 2002;16(5):761–75.
7. Traeger AC, et al. Care for low back pain: can health systems deliver? Bull World Health Organ. 2019;97(6):423–33.
8. Kinge JM, et al. Musculoskeletal disorders in Norway: prevalence of chronicity and use of primary and specialist health care services. BMC Musculoskelet Disord. 2015;16(1):1–9.
9. Beyera GK, O'Brien J, Campbell S. Health-care utilisation for low back pain: a systematic review and meta-analysis of population-based observational studies. Rheumatol Int. 2019;39(10):1663–79.
10. Licciardone JC. The epidemiology and medical management of low back pain during ambulatory medical care visits in the United States. Osteopath Med Prim Care. 2008;2:11.
11. Foster NE, et al. Prevention and treatment of low back pain: evidence, challenges, and promising directions. Lancet. 2018.
12. Vlaeyen JWS, et al. Low back pain. Nat Rev Dis Primers. 2018;4(1):52.
13. Riley RD, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. PLoS Med. 2013;10(2):e1001380.
14. Mutubuki EN, et al. Predictive factors of high societal costs among chronic low back pain patients. Eur J Pain. 2019.
15. Becker A, et al. Low back pain in primary care: costs of care and prediction of future health care utilization. Spine (Phila Pa 1976). 2010;35(18):1714–20.
16. Engel CC, von Korff M, Katon WJ. Back pain in primary care: predictors of high health-care costs. Pain. 1996;65(2-3):197–204.
17. Ritzwoller DP, et al. The association of comorbidities, utilization and costs for patients identified with low back pain. BMC Musculoskelet Disord. 2006;7:72.
18. Wammes JJG, et al. Systematic review of high-cost patients' characteristics and healthcare utilisation. BMJ Open. 2018;8(9).
19. Stewart WF, et al. Patterns of health care utilization for low back pain. J Pain Res. 2015;8:523–35.
20. Paeck T, et al. Are older adults missing from low back pain clinical trials? A systematic review and meta-analysis. Arthritis Care Res. 2014;66(8):1220–6.
21. Joud A, Petersson IF, Englund M. Low back pain: epidemiology of consultations. Arthritis Care Res. 2012;64(7):1084–8.
22. Hemingway H, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. BMJ. 2013;346:e5595.
23. Killingmo, R.M., et al., Statistical analysis plan (SAP) for: Modifiable prognostic factors of high cost related to healthcare utilization among older people seeking primary care with a new episode of back pain - an identification and replication study. 2021, ClinicalTrials.gov: https://clinicaltrials.gov/ProvidedDocs/09/NCT04261309/SAP_002.pdf.
24. McShane LM, et al. REporting recommendations for tumour MARKer prognostic studies (REMARK). Br J Cancer. 2005;93(4):387–91.
25. Grotle, M., BACk Pain in Elders in Norway (BACE-N). 2020, ClinicalTrials.gov: https://clinicaltrials.gov/ct2/show/NCT04261309.
26. Scheele J, et al. Back complaints in the elders (BACE); design of cohort studies in primary care: an international consortium. BMC Musculoskelet Disord. 2011;12:193.
27. Kanters TA, et al. Update of the Dutch manual for costing studies in health care. PLoS One. 2017;12(11):e0187477.
28. Wenig CM, et al. Costs of back pain in Germany. Eur J Pain. 2009;13(3):280–6.
29. Ferreira M, et al. Factors defining care-seeking in low back pain - A meta-analysis of population based surveys. Eur J Pain. 2010;14(7):747.e1–7.
30. Lentz TA, et al. Factors associated with persistently high-cost health care utilization for musculoskeletal pain. PLoS One. 2019;14(11):e0225125.
31. Von Korff M, Jensen MP, Karoly P. Assessing global pain severity by self-report in clinical and health services research. Spine (Phila Pa 1976). 2000;25(24):3140–51.
32. Lim KL, Jacobs P, Klarenbach S. A population-based analysis of healthcare utilization of persons with back disorders: results from the Canadian Community Health Survey 2000-2001. Spine (Phila Pa 1976). 2006;31(2):212–8.
33. Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. Spine (Phila Pa 1976). 1983;8(2):141–4.
34. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992;30(6):473–83.
35. Keeley P, et al. Psychosocial predictors of health-related quality of life and health service utilisation in people with chronic low back pain. Pain. 2008;135(1-2):142–50.
36. Radloff LS. The CES-D scale: a Self-Report Depression Scale for Research in the General Population. Appl Psychol Meas. 1977;1(3):385–55.
37. Waddell G, et al. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain. 1993;52(2):157–68.
38. Sangha O, et al. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. Arthritis Rheum. 2003;49(2):156–63.
39. Rattay P, et al. Utilization of outpatient and inpatient health services in Germany: results of the German Health Interview and Examination Survey for Adults (DEGS1). Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz. 2013;56(5-6):832–44.
40. Chechulin Y, et al. Predicting patients with high risk of becoming high-cost healthcare users in Ontario (Canada). Healthc Policy. 2014;9(3):68–79.
41. Hoebel J, et al. Socioeconomic Status and Use of Outpatient Medical Care: The Case of Germany. PLoS One. 2016;11(5):e0155982.
42. Fritz JM, Kim J, Dorius J. Importance of the type of provider seen to begin health care for a new episode low back pain: associations with future utilization and costs. J Eval Clin Pract. 2016;22(2):247–52.

43. Box GE, Tidwell PW. Transformation of the independent variables. Technometrics. 1962;4:531–50.
44. Kent P, et al. A conceptual framework for prognostic research. BMC Med Res Methodol. 2020;20(1):172.
45. Harrell FE Jr, et al. Regression modelling strategies for improved prognostic prediction. Stat Med. 1984;3(2):143–52.
46. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361–87.
47. Steyerberg EW, et al. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. Med Decis Mak. 2001;21(1):45–56.
48. Steyerberg EW, et al. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med. 2000;19(8):1059–79.
49. Ambler G, Brady AR, Royston P. Simplifying a prognostic model: a simulation study based on clinical data. Stat Med. 2002;21(24):3803–22.
50. Moons KG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014;11(10):e1001744.
51. Pavlou M, et al. Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. Stat Med. 2016;35(7):1159–77.
52. Cowley LE, et al. Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature. Diagn Progn Res. 2019;3:16.
53. Pavlou M, et al. How to develop a more accurate risk prediction model when there are few events. BMJ. 2015;351:h3868.
54. Budtz CR, Mose S, Christiansen DH. Socio-demographic, clinical and psychological predictors of healthcare utilization among patients with musculoskeletal disorders: a prospective cohort study. BMC Health Serv Res. 2020;20(1):239.
55. Lentz TA, Beneciuk JM, George SZ. Prediction of healthcare utilization following an episode of physical therapy for musculoskeletal pain. BMC Health Serv Res. 2018;18.
56. Franklin M, et al. An Educational Review About Using Cost Data for the Purpose of Cost-Effectiveness Analysis. Pharmacoeconomics. 2019;37(5):631–43.
57. Leurent B, Gomes M, Carpenter JR. Missing data in trial-based cost-effectiveness analysis: An incomplete journey. Health Econ. 2018;27(6):1024–40.
58. Briggs A, et al. Missing… presumed at random: cost-analysis of incomplete data. Health Econ. 2003;12(5):377–92.
59. Petrou S, et al. The accuracy of self-reported healthcare resource utilization in health economic studies. Int J Technol Assess Health Care. 2002;18(3):705–10.
60. Icks A, et al. Agreement found between self-reported and health insurance data on physician visits comparing different recall lengths. J Clin Epidemiol. 2017;82:167–72.
61. Short ME, et al. How accurate are self-reports? Analysis of self-reported health care utilization and absence when compared with administrative data. J Occup Environ Med. 2009;51(7):786–96.
62. Hunger M, et al. Official statistics and claims data records indicate non-response and recall bias within survey-based estimates of health care utilization in the older population. BMC Health Serv Res. 2013;13:1.
63. Andersen RM. National health surveys and the behavioral model of health services use. Med Care. 2008;46(7):647–53.
64. Slagsvold B, Veenstra M, SO D, Hagestad G, Hansen T, Herlofson K, et al. Life-course, ageing and generations in Norway:the NorLAG study. Norsk Epidemiologi. 2012;22(2).
65. Vigdal ON, et al. Characteristics of older adults with back pain associated with choice of first primary care provider: a cross-sectional analysis from the BACE-N cohort study. BMJ Open. 2021;11(9):e053229.
66. Grotle M, et al. Prognostic factors in first-time care seekers due to acute low back pain. Eur J Pain. 2007;11(3):290–8.
67. Nordstoga AL, et al. Improvement in Work Ability, Psychological Distress and Pain Sites in Relation to Low Back Pain Prognosis A Longitudinal Observational Study in Primary Care. Spine. 2019;44(7):E423–9.
68. Buchbinder R, et al. The Lancet Series call to action to reduce low value care for low back pain: an update. Pain. 2020;161(Suppl 1):S57–64.

Publisher's Note

Paper III

Munk R, Storheim K, Småstuen MC, Grotle M

Measuring Productivity Costs in Patients With Musculoskeletal Disorders: Measurement Properties of the Institute for Medical Technology Assessment Productivity Cost Questionnaire

Value in Health. 2019; 22(12):1410-1416. DOI: 10.1016/j.jval.2019.07.011

ScienceDirect

Contents lists available at **sciencedirect.com**
Journal homepage: **www.elsevier.com/locate/jval**

Methodology

Measuring Productivity Costs in Patients With Musculoskeletal Disorders: Measurement Properties of the Institute for Medical Technology Assessment Productivity Cost Questionnaire

Rikke Munk, MSc,[1,*] Kjersti Storheim, PhD,[1,2] Milada C. Småstuen, PhD,[1] Margreth Grotle, PhD[1,2]

[1]*Department of Physiotherapy, Oslo Metropolitan University, Oslo, Norway;* [2]*Division of Clinical Neuroscience, Research and Communication Unit for Musculoskeletal Health, Oslo University Hospital, Oslo, Norway*

ABSTRACT

Background: The Institute for Medical Technology Assessment Productivity Cost Questionnaire (iPCQ) was recently developed to cover all domains of productivity costs; absenteeism, presenteeism and productivity costs related to unpaid work. The original iPCQ has not been tested with respect to neither content or construct validity, nor reliability, and there is no Norwegian version of the questionnaire.

Objectives: To translate and cross-culturally adapt the iPCQ into Norwegian and to test its measurement properties among patients with musculoskeletal disorders.

Methods: Translation and cross-cultural adaptation was conducted according to guidelines, and measurement properties were investigated using a cross-sectional design including a test–retest assessment. Patients with musculoskeletal disorders were recruited from secondary care. Data quality, content validity (10 patients evaluated comprehensibility, 2 researchers and 1 clinician evaluated relevance and comprehensiveness), construct validity (factor analysis, internal consistency, divergent hypothesis testing), and test–retest reliability (intraclass correlation coefficient two-way random average agreement, Cohen's unweighted kappa) were assessed.

Results: In total, 115 patients with a mean age (SD) of 46 (9) years were included, and 62 responded to the retest. The questionnaire was feasible, with little missing data and no floor or ceiling effects. Content validity displayed good comprehensibility and relevance and sufficient comprehensiveness. Factor analysis revealed a 3-component solution accounting for 82% of the total variance; items loaded as expected and supported the original structure of the iPCQ. Internal consistency was acceptable for the 3 components of productivity cost, with an inter-item correlation ranging from 0.42 to 0.62. Further, a total of 91% of our hypotheses were verified. The intraclass correlation coefficient values ranged from 0.88 to 0.99 for all items except one; kappa ranged from 0.61 to 0.92, indicating overall good reliability of the questionnaire.

Conclusions: The Norwegian iPCQ showed good measurement properties among patients with musculoskeletal disorders from secondary care in Norway. We therefore recommend the iPCQ as a useful tool for measuring productivity costs in patients with musculoskeletal disorders.

Keywords: measurement properties, musculoskeletal disorders, productivity costs.

VALUE HEALTH. 2019; 22(12):1410–1416

Introduction

Musculoskeletal disorder is one of the leading causes of disability worldwide,[1] accounting for a huge amount of productivity loss.[2] The impact of disease and disorders on productivity is an important part of health economic evaluations. When a societal perspective is included in research, it can provide information on the relative cost of different disorders and on the relative cost-effectiveness of healthcare interventions, and it is therefore an important tool in decision making for how to best allocate resources.[3,4] Currently, there is no gold standard for measuring productivity costs.[5-7] Nevertheless, there is a general agreement that one should measure not only the productivity costs related to absence from paid work (absenteeism) and reduced productivity while at paid work (presenteeism) but also costs related to unpaid work such as household work, care work, and volunteer work.[4] The Institute for Medical Technology Assessment (iMTA) Productivity Cost Questionnaire (iPCQ) was recently developed to cover these 3 domains of productivity costs. It was designed to capture core parts of existing questionnaires and to be a short, generic, patient-reported outcome

* Address correspondence to: Rikke Munk, MSc, Department of Physiotherapy, Oslo Metropolitan University, Pilestredet 50, 0167 Oslo, Norway. Email: rikke@oslomet.no

https://doi.org/10.1016/j.jval.2019.07.011

measure, allowing for quantification and valuation of all productivity costs.[4] Two studies have tested some of the measurement properties of the iPCQ. Bouwmans et al[4] confirmed its feasibility and face validity. In a modified version (iPCQ-VR), Beemster et al[8] tested reliability, agreement, and responsiveness of the core parts of absenteeism and presenteeism; they found good measurement properties on long-term sick leave and poor measurement properties on short-term sick leave and presenteeism. To the best of our knowledge, the original iPCQ version has not been tested with respect either to content or construct validity or to the reliability of the entire questionnaire. Furthermore, there is no Norwegian version of this instrument. Therefore, the purpose of this study was to translate and cross-culturally adapt the original iPCQ into Norwegian and to test its measurement properties among patients with musculoskeletal disorders.

Methods

Design

This study was carried out in 2 stages. First, the original version of the iPCQ was translated and cross-culturally adapted into Norwegian. The Norwegian iPCQ was then tested for its measurement properties using a cross-sectional design. In addition, a test–retest assessment was conducted after 2 to 3 days.

Translation and Cross-Cultural Adaptation

The translation and cross-cultural adaptation was carried out according to international guidelines.[9,10] Two persons (1 philologist and 1 clinician), whose mother tongue was Norwegian, independently translated the original iPCQ from English into Norwegian. The 2 Norwegian versions were then synthesized into 1 version before being translated back into English. Two native English speakers (1 philologist and 1 clinician), both blinded to the original iPCQ, independently performed the back-translation and synthesized the 2 English versions into one. An expert committee consisting of the translators and 2 researchers in our research group reviewed all translations. In a formal meeting, the committee discussed deviations until consensus on a prefinal version was reached. The goal was for the prefinal Norwegian iPCQ to be as concise and easy to understand as possible. The prefinal version was tested on 10 patients with musculoskeletal disorders. None of the patients had difficulty understanding the meaning of items or responses, and they found it easy to comprehend. No changes had to be made, so the final version of the Norwegian iPCQ evaluated in this study is the same as the prefinal version.

Participants

We planned to recruit 100 patients based on quality criteria recommended by Terwee et al[11] and Nunnally.[12] These criteria suggest a minimum of 100 participants for assessing internal consistency, at least 50 participants for assessing reproducibility and floor or ceiling effects,[11] and at least 10 participants for each item being included in the factor analysis.[12]

Participants were recruited from secondary care at an outpatient rehabilitation clinic in Akershus, Norway, between November 2015 and January 2018. Eligible participants were patients with musculoskeletal disorders, aged 18 years or older, who were working or on sick leave. The exclusion criterion was the inability to speak, read, or write in Norwegian. Inclusion was performed by a clinician, primarily a physiotherapist, who met the patients at the clinic. At baseline, all patients received written and oral information about the study, and signed informed consent was obtained from all patients. The study was classified as a quality assessment study by the Norwegian Regional Committee for Medical Research Ethics (reference No. 2014/1634/REK vest) and was approved by the Norwegian Social Science Data Service (reference No. 45367) in 2015.

Procedures and Measurements

At baseline, the included patients completed the iPCQ as part of a comprehensive questionnaire, which also included sociodemographic variables, pain localization, pain intensity and history, health-related quality of life, physical workload at work, and psychosocial work environment. The McGill pain drawing and the Numeric Rating Scale were used to measure pain localization and intensity.[13,14] The Short Form–36 Health Status Questionnaire was used to measure health-related quality of life,[15] the Physical Workload Questionnaire was used to measure physical workload at work,[16] and questions from the QPS Nordic questionnaire were used to measure characteristics of the psychosocial work environment.[17]

Patients consenting to participate in the reproducibility part of the study completed the iPCQ at their second attendance, preferably with a 2- to 3-day interval.

The iMTA Productivity Cost Questionnaire

The iPCQ consists of 18 items and adopts a recall period of 4 weeks. In the introduction, 9 items (numbers A1 to A6 and 1 to 3) assess the date for reply and the following sociodemographic factors: age, sex, education level, work status, paid or unpaid work, profession, number of workdays, and work hours per week of paid work. Further, productivity costs are measured in 3 separate index scores with individual sum scores: absence from paid work (absenteeism), reduced productivity at paid work (presenteeism), and productivity loss in unpaid work. To calculate productivity costs, 8 core items are used. The value of absenteeism is calculated from items 2, 3, 4, and 6; presenteeism from items 2, 3, 8, and 9; and unpaid work productivity loss from items 11 and 12.[18] The costs of productivity loss are valued in hours; hence, they can be translated by a standard cost price of productivity per hour.

The 3 items (items 4 to 6) measuring productivity costs due to short- and long-term absence from paid work originate from the PRODISQ[19] and the SF-HLQ[20] and identify the occurrence and length of absenteeism. The validity of these questions in terms of feasibility, reliability, and construct validity (comparison between long-term absence and register data) has been demonstrated in previous studies.[21,22]

The 3 items (items 7 to 9) measuring productivity costs owing to presenteeism at paid work are composed of questions from the PRODISQ and the SF-HLQ and aim to identify whether the responders suffered from health problems at work and, if so, for how many days. As well, the responders are asked to rate their work performance on days with productivity loss in comparison with function on normal working days using an 11-point rating scale. The reliability of these questions has been shown to be acceptable using a test–retest design.[21]

The items (items 10 to 12) about productivity costs from unpaid work were developed at the iMTA at Erasmus University Rotterdam. The responders are asked whether they can perform less unpaid work, such as volunteer work and household work, as a result of health problems, and to state how many hours it would take someone else to replace this unperformed work.[21]

The English and the Norwegian versions and the manual for the iPCQ are available from the iMTA at Erasmus University Rotterdam.[18,23]

Analysis

All data analyses were performed using SPSS version 24 (SPSS Inc, Chicago, IL), and the Vassarstats kappa was calculated using

Table 1. Patient demographic characteristics and clinical status.

	Validity study (n = 115)		Test-retest study (n = 62)	
		Missing, n (%)		Missing, n (%)
Male, n (%)	36 (31.3)	0 (0)	23 (37.1)	0 (0)
Age in years, mean (SD)	45.6 (9.3)	0 (0)	46.3 (8.5)	0 (0)
Education level high, n (%)	67 (58.2)	0 (0)	35 (56.5)	0 (0)
Mother tongue Norwegian, n (%)	100 (87.0)	0 (0)	53 (85.5)	0 (0)
Work status, n (%)	—	0 (0)	—	0 (0)
Employed or self-employed (paid job)	104 (90.4)	0 (0)	55 (88.7)	0 (0)
Sick leave during past 4 weeks	79 (68.7)	1 (0.9)	40 (64.5)	0 (0)
Rehabilitation, work disability	7 (6.1)	0 (0)	4 (6.5)	0 (0)
Pain period in days, median (range)	700 (5-10 950)	26 (22.6)	720 (5-10 950)	15 (13.0)
Pain severity last week (NRS 0-10), median (range)	5 (1-9)	3 (2.6)	5 (1-9)	1 (1.6)
Pain location, n (%)	—	0 (0)	—	0 (0)
Lower limbs	70 (60.9)	—	36 (58.1)	—
Back	80 (69.6)	—	47 (75.8)	—
Neck	57 (49.6)	—	26 (41.9)	—
Upper limbs	55 (47.8)	—	24 (38.7)	—
>2 pain areas	45 (39.1)	—	23 (37.1)	—
Health-related quality of life (SF-36 0-100), median (range)	—	—	—	—
Mental health	70.0 (10-100)	1 (0.9)	75.0 (10-95)	1 (1.6)
Physical function	75 (30-100)	0 (0)	70 (30-95)	0 (0)
Physical workload (PWQ 0-100), median (range)	—	—	—	—
Heavy physical workload	20.8 (0-86.1)	7 (6.1)	26.4 (0-86.1)	4 (6.5)
Long-lasting posture and repetitive movement	50.0 (0-100)	3 (4.8)	50.0 (0-100)	3 (4.8)
QPS Nordic (1-5), median (range)				
Control of decisions	2.8 (1-5)	7 (6.1)	2.8 (1-5)	3 (4.8)
Authorizing management	3.3 (1-5)	7 (6.1)	3.7 (1-5)	6 (9.7)
Role conflict	2.3 (1-5)	5 (4.4)	2.3 (1-5)	4 (6.5)
Fair leadership	4.0 (1-5)	13 (11.3)	4.0 (1.3-5)	9 (14.5)

NRS indicates Numeric Rating Scale; PWQ, Physical Workload Questionnaire; QPS Nordic, General Nordic questionnaire for psychological and social factors at work; SF-36, 36-Item Short-Form Health Survey.

http://vassarstats.net/kappa.html. The measurement properties of the Norwegian iPCQ were tested as follows.

Data quality

Proportions of missing data and floor or ceiling effects were described. Floor or ceiling effects were considered to be present if greater than 15% of participants reported the lowest or highest possible score.[11]

Content validity

To assess content validity, the COSMIN group recommends evaluating relevance, comprehensiveness, and comprehensibility.[24] In the present study, we asked, with open questions, 10 patients with musculoskeletal disorders about the comprehensibility of the iPCQ (are the instructions, items, and responses understood as intended; are the items appropriately worded; and do the response options match the question?). Two researchers and 1 clinician, with no conflict of interest, were asked about the relevance and the comprehensiveness of the iPCQ (are all included items relevant for the construct of productivity cost and the target population; are all key elements of productivity costs included?).

Construct validity

To assess construct validity, the COSMIN group recommends evaluating structural validity and internal consistency, followed by hypothesis testing.[24,25]

In the present study, we expected the iPCQ to cover the 3 components of productivity costs: absenteeism, presenteeism, and unpaid work productivity costs. To confirm the underlying structure of the questionnaire and to investigate structural validity of the iPCQ, confirmatory factor analysis was conducted.[26] A computed factor loading expresses the strength or magnitude of an association between a given item and a factor. The loading ranges from 0 to 1; the higher the value, the more an item is associated with a factor (a component).[26] Based on the original structure of the questionnaire, we expected that the core items would load onto 3 components: absenteeism (items 2, 3, 4, and 6), presenteeism (items 2, 3, 8, and 9), and unpaid work productivity costs (items 11 and 12). Furthermore, we hypothesized that the internal consistency was sufficient for all 3 components of the iPCQ. Internal consistency of the components was assessed using interitem correlation. For the components to be considered sufficiently reliable, the interitem correlation should be greater than 0.4.[27]

Finally, we hypothesized that high productivity costs, assessed with the 3 index scores of the iPCQ, were negatively correlated with low health-related quality of life[28,29] and positively correlated with low physical workload,[19,30-32] low psychosocial work environment,[30-35] and much pain[28,30,33,36,37] (divergent construct validity). These hypotheses were based on previous studies.[19,28-37] In general, there is some inconsistency in the literature, but to the best of our knowledge, it appears that most available studies

Table 2. Descriptive statistics, including missing data, for the iPCQ core items and the index scores (n = 115).*

Core items (label, wording, and response format)	Missing, n (%)	NA, n (%)	Median (range)
#2. Weekly work hours *How many hours a week do you work? Count only the hours that you get paid (. . . hours)*	3 (2.6)	11 (9.6)	37.5 (7.5-52)
#3. Weekly workdays *How many days a week do you work? (. . . days)*	2 (1.7)	11 (9.6)	5 (2-7)
#4. Number of days absenteeism short term *Have you missed work in the last 4 weeks as a result of being sick? (No; Yes I have missed . . . days)*	5 (4.3)	35 (30.4)	14 (1-21)
#6. Numbers of days absenteeism long term *Did you miss work earlier than the period of 4 weeks due to being sick? This is referring to one whole uninterrupted period of missed work as a result of being sick. (No, Yes). If yes, when did you call in sick? (day, month, year)*	3 (2.6)	46 (40.0)	175 (18-586)
#8. Number of workdays with disability *How many days at work were you bothered by physical or psychological problems? Only count the days at work in the last 4 weeks. (. . . workdays)*	0 (0)	83 (72.2)	11 (3-28)
#9. Effective score completed work *On the days that you were bothered by these problems, was it perhaps difficult to get as much work finished as you normally do? On these days how much work could you do on average?* *Look at the figures below. A 10 means that you were able to do as much work as you normally do. A 0 means that you were unable to do any work on these days. (0-10, Likert-type scale)*	0 (0)	83 (72.2)	7 (3-10)
#11. Number of days less unpaid work *How many days did this happen? Only count the days in the last 4 weeks. (…. days)*	7 (6.1)	59 (51.3)	15 (1-28)
#12. Number of hours less unpaid work *Imagine that somebody, for example, your partner, family member, or friend, helped you on these days, and he or she did all the unpaid work that you were unable to do for you. How many hours on average did that person spend doing this on these days? (On average . . . hours on these days)*	11 (9.6)	59 (51.3)	20 (2-280)
Index scores			
Absenteeism, hours	9 (7.8)	35 (30.4)	1155 (8-4688)
Presenteeism, hours	1 (0.9)	86 (74.8)	25 (5-105)
Productivity loss unpaid work, hours	12 (10.4)	59 (51.3)	327 (6-7840)

iPCQ indicates Institute for Medical Technology Assessment Productivity Cost Questionnaire.
*The index score of absenteeism is calculated from core items 2, 3, 4, and 6; presenteeism from core items 2, 3, 8, and 9; and productivity loss unpaid work from core items 11 and 12.[18] NA indicates not applicable, due to the structure of the iPCQ.

demonstrate a low correlation between these variables and productivity costs. Spearman's rho was used in all correlation analyses because the scales were not normally distributed. Correlation coefficients less than 0.3, between 0.3 and 0.6, and greater than 0.6 were considered low, moderate, and high, respectively.[38]

Reliability

The test–retest reliability of the continuous variables (items 2, 3, 4 second part, 6, 8, 9, 11, and 12) and the 3 index scores was assessed with the intraclass correlation coefficient (ICC) using 2-way random, average agreement. The acceptable level of ICC was set to >0.70.[11] In addition, Cohen's unweighted kappa was used for dichotomous variables (items 4 first part, 5, 7, and 10) of the iPCQ. Kappa values were categorized according to Altman: poor (0 to 0.2), fair (0.21 to 0.40), moderate (0.41 to 0.60), good (0.61 to 0.80), and very good (0.81 to 1.00).[39]

Results

A total of 115 patients with a mean age (SD) of 46 (9) years were included in the cross-sectional study, and a sample of 62 participants completed the retest questionnaire. The median time interval between test and retest was 3 days (range, 1-10 days). Almost all included patients were in paid work (90%), and more than half had been on sick leave during the previous 4 weeks. On average, they reported moderate pain, and the most frequently reported pain area was the back region. Study sample characteristics are shown in Table 1.

Data quality

The proportion of missing data was relatively small: less than 10% for all items (ranging from 0% to 9.6%). There were no floor or ceiling effects for any of the items. All continuous variables in the iPCQ and the sum scores for the 3 index scores had a skewed

Table 3. Confirmatory factor analyses with item loading.

	Component		
Item	1	2	3
Weekly work hours	0.77	0.55	
Weekly workdays	0.77	0.57	
Numbers of days absenteeism short term	0.81		
Numbers of days absenteeism long term	0.70	−0.39	
Numbers of workdays with disability		0.82	0.34
Effective score completed work		0.85	
Number of days less unpaid work			0.89
Number of hours less unpaid work			0.92

Factor loading greater than 0.3 is reported.

Table 4. Correlation of iPCQ domains with other health-related variables.

iPCQ domain	n	Correlation coefficient* (*P* value)
Absenteeism		
Physical function health related QOL (SF-36)	95	−0.107 (.301)
Mental-health-related QOL (SF-36)	94	−0.210 (.042)
Heavy physical workload (PWQ)	95	0.157 (.128)
Long-lasting posture and repetitive movement (PWQ)	95	0.165 (.110)
Pain intensity last week (NRS)	92	0.194 (.064)
Psychosocial work environment (QPS Nordic)		
Control of decisions	89	−0.336 (.001)
Authorizing leadership	92	−0.087 (.409)
Role conflict	93	0.035 (.736)
Fair leadership	87	0.212 (.048)
Presenteeism		
Physical function health-related QOL (SF-36)	43	−0.058 (.713)
Mental-health-related QOL (SF-36)	43	−0.202 (.194)
Heavy physical workload (PWQ)	43	0.341 (.025)
Long-lasting posture and repetitive movement (PWQ)	43	0.113 (.472)
Pain intensity last week (NRS)	42	0.133 (.399)
Psychosocial work environment (QPS Nordic)		
Control of decisions	40	−0.033 (.840)
Authorizing management	43	0.054 (.733)
Role conflict	43	0.297 (.053)
Fair leadership	39	−0.239 (.143)
Unpaid work productivity loss		
Physical function health-related QOL (SF-36)	103	−0.182 (.067)
Mental-health-related QOL (SF-36)	103	−0.179 (.071)
Pain intensity last week (NRS)	100	0.195 (.052)
Number of pain locations (McGill pain drawing)	103	0.205 (.038)

iPCQ indicates Institute for Medical Technology Assessment Productivity Cost Questionnaire; NRS, Numeric Rating Scale; PWQ, Physical Workload Questionnaire; QPS Nordic, General Nordic questionnaire for psychological and social factors at work; SF-36, 36-Item Short-Form Health Survey.
*Calculated as Spearman's rho.

distribution. Descriptive statistics for the core items and the index scores are listed in Table 2.

Content validity

Overall, the iPCQ has sufficient content validity. The included items are relevant and cover all domains of productivity costs, except compensation mechanisms and part-time sick leave. Moreover, the questionnaire was understood as intended by the 10 patients evaluating comprehensibility. Nevertheless, when going through all responses in the validity study, we found some deviations in the question about "number of hours with less unpaid work." Some patients recorded hours per day, whereas others recorded the total number of hours in the 4-week period.

Construct validity

The confirmatory factor analysis revealed a 3-component solution accounting for 82% of the total variance in the data. The first, second, and third components explained 31%, 29%, and 23% of the total variance, respectively. Furthermore, the confirmatory factor analysis displayed, as expected, that core items 2, 3, 4, and 6 load on component 1; that core items 2, 3, 8, and 9 load on component 2; and that core items 11 and 12 load on component 3 (Table 3).

The internal consistency and the level of interitem correlation were acceptable for the 3 components with values of 0.46, 0.42, and 0.62 for absenteeism, presenteeism, and productivity costs from unpaid work, respectively.

The results for the divergent construct validity are presented in Table 4. As expected, the correlation coefficients have low values (<0.30) between the 3 index scores of the iPCQ and almost all of the health- and work-related variables. There was a moderate correlation between absenteeism and control of decisions and between presenteeism and heavy physical workload. A total of 91% of our hypotheses were confirmed.

Reliability

Table 5 shows the test–retest reliability results for the 3 index scores and the core items. According to the ICCs, reliability was greater than the recommended minimum standard for the 3 index scores and all core items, with the exception of item 8 (number of workdays with disability; 0.34). The ICCs for the iPCQ index scores ranged from 0.89 to 0.99. Kappa values of the 4 dichotomous items of iPCQ ranged from 0.62 to 0.84. Item 7 for work despite disability during the last 4 weeks scored the lowest, and item 5 for absenteeism before the last 4 weeks scored the highest.

Discussion

In this study, we assessed the measurement properties of the Norwegian version of the iPCQ, and the overall results were good with respect to data quality, content and construct validity, and reliability.

The iPCQ is a relatively new questionnaire, and there are only 2 previous studies with which we can compare our results.[4,8]

The data quality in our study, showing little missing data, is in line with the original study of Bouwmans et al,[4] in which the iPCQ was tested in the general population.

Our evaluation of content validity showed good comprehensibility of the iPCQ, with the exception of item 12 (number of hours

Table 5. Test–retest reliability results for the iPCQ core items and the index scores.

Core items (label) and index scores (n)	Median test (interquartile range)	Median retest (interquartile range)	% Agreement	ICC (95% CI)
Weekly work hours (48)	37.5 (37.5-40)	37.5 (37.5-40)	91.7	0.92 (0.86-0.96)
Weekly workdays (52)	5 (5-5)	5 (5-5)	96.2	0.88 (0.79-0.93)
Number of days absenteeism short term (34)	13.5 (12-20)	12 (12-20)	79.4	0.92 (0.83-0.96)
Number of days absenteeism long term (29)	197 (86-312)	199 (78-315)	86.2	0.99 (0.99-0.99)
Number of workdays with disability (18)	16 (8-20)	19 (10-20)	72.2	0.34 (−0.89-0.76)
Effective score completed work (20)	7 (5-8)	7 (5-8)	75.0	0.98 (0.94-0.99)
Number of days less unpaid work (21)	15 (7-23)	16 (5-24)	52.4	0.88 (0.69-0.95)
Number of hours less unpaid work (20)	20 (9-29)	11 (9-24)	85.0	0.99 (0.99-0.99)
Index scores				
Absenteeism long term, hours (23)	1576 (698-2363)	1592 (555-2393)	—	0.99 (0.98-0.99)
Absenteeism short term, hours (2)	34.3 (16-53)	53.5 (32-75)	—	0.89 (−0.04-1.0)
Presenteeism, hours (16)	23 (10-49)	31 (14-48)	—	0.91 (0.74-0.97)
Productivity loss unpaid work, hours (19)	294 (48-500)	80 (50-560)	—	0.98 (0.94-0.99)

% agreement indicates (number of identical/total answers) × 100; CI, confidence interval; ICC, intraclass correlation coefficient agreement, 2-way random, average measures; iPCQ, Institute for Medical Technology Assessment Productivity Cost Questionnaire.

with less unpaid work). Some patients recorded hours per day, whereas others recorded the total number of hours in the 4-week period. Hence, a small adjustment was made to the questionnaire after finishing this study, and participants are now specifically requested to record hours per day. Further, the included items were considered to be relevant and to cover the main domains of productivity costs. Nevertheless, there are 2 exceptions: The iPCQ does not cover compensation mechanisms or part-time sick leave. Compensation mechanisms may influence the total value of productivity costs.[4,5,40] Nonetheless, the extent to which these mechanisms affect final productivity costs remains unclear, and adjusting for them currently seems premature.[4,5,40] Nevertheless, it would be possible to include items covering part-time sick leave, thereby increasing the usefulness of the iPCQ. In the iPCQ-VR, Beemster et al[8] showed that part-time sick leave could be reliably measured in patients with nonspecific musculoskeletal pain.

Our assessment of construct validity confirmed a 3-component solution, which is similar to the original study of the iPCQ.[4] Bouwmans et al[4] distinguished between the 3 components based on a theoretical rationale. Our study is the first to confirm that the core items load as expected and that the 3 components accounted for as much as 82% of the total variance in the data. The internal consistency was acceptable for the 3 components. Furthermore, construct validity was supported by the hypothesis testing, which has not been reported previously.

In the present study, the iPCQ showed good reliability for the index scores and all of the individual items, except item 8 (number of workdays with disability). The study of Beemster et al[8] supports a good reliability of the items related to long-term sick leave and a low reliability of the item covering number of workdays with disability. Regarding items related to short-term sick leave (item 4) and presenteeism (item 9), our results indicated a higher reliability than was found by Beemster et al.[8] This difference might be explained by a different time interval between test and retest, as Beemster et al[8] used an average of 20 days compared with 3 in this study. The reliability of items 11 and 12, covering productivity costs from unpaid work, have not been tested previously.

The main limitation of this study is that we did not compare the index scores for absenteeism against public register data, which is recommended.[40] We are, however, conducting a new study in which this comparison will be carried out. A second potential weakness is the hypothesis testing in the present study. Because construct validity is concerned with how well an instrument captures the intended construct,[6,11] it is questionable to what degree divergent hypothesis testing alone can be used to provide evidence for construct validity. We could have tested the iPCQ against questionnaires that measure the same construct; however, there is no gold standard regarding the measurement of productivity cost, and different questionnaires often result in varied estimates, especially of presenteeism.[6,20,22,41-43] A third potential weakness of this study is that we have tested measurement properties of a generic questionnaire only in a sample of patients with musculoskeletal disorders in Norway. Obviously, testing measurement properties in different settings and including patients with different disorders is recommended. A fourth weakness of this study is the lack of data on eligible study participants who declined to participate. Because of limited resources, it was not possible to record information on all patients attending the rehabilitation clinic during the data collection period.

The main strength of the present study is that it is the first to test the content and construct validity as well as reliability of the original iPCQ and that this testing was conducted in line with COSMIN guidelines.[11]

Conclusion

This study showed that the iPCQ has good measurement properties for measuring productivity costs among patients with musculoskeletal disorders receiving rehabilitation in a secondary outpatient clinic in Norway. We can recommend using the Norwegian version of the iPCQ for clinical and research purposes on patients with musculoskeletal disorders. Because the iPCQ is a generic instrument, further studies should validate it in other patient populations.

Acknowledgments

Ingeborg Landstad and Unicare Friskvern contributed to data collection.

Source of financial support: This study was supported by Oslo Metropolitan University.

REFERENCES

1. Global Burden of Disease Study Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2015;386(9995):743–800.
2. Bevan S. Economic impact of musculoskeletal disorders (MSDs) on work in Europe. *Best Pract Res Clin Rheumatol*. 2015;29(3):356–373.
3. Burdorf A. Economic evaluation in occupational health–its goals, challenges, and opportunities. *Scand J Work Environ Health*. 2007;33(3):161–164.
4. Bouwmans C, Krol M, Severens H, et al. The iMTA Productivity Cost Questionnaire: a standardized instrument for measuring and valuing health-related productivity losses. *Value Health*. 2015;18(6):753–758.
5. Krol M, Brouwer W, Rutten F. Productivity costs in economic evaluations: past, present, future. *Pharmacoeconomics*. 2013;31(7):537–549.
6. Tang K. Estimating productivity costs in health economic evaluations: a review of instruments and psychometric evidence. *Pharmacoeconomics*. 2015;33(1):31–48.
7. Zhang W, Bansback N, Anis AH. Measuring and valuing productivity loss due to poor health: a critical review. *Soc Sci Med*. 2011;72(2):185–192.
8. Beemster TT, van Velzen JM, van Bennekom CAM, Reneman MF, Frings-Dresen MHW. Test-retest reliability, agreement and responsiveness of productivity loss (iPCQ-VR) and healthcare utilization (TiCP-VR) questionnaires for sick workers with chronic musculoskeletal pain. *J Occup Rehabil*. 2019;29(1):91–103.
9. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol*. 1993;46(12):1417–1432.
10. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*. 2000;25(24):3186–3191.
11. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34–42.
12. Nunnally JO. *Psychometric Theory*. New York, NY: McGraw-Hill; 1978.
13. Escalante A, Lichtenstein MJ, White K, Rios N, Hazuda HP. A method for scoring the pain map of the McGill Pain Questionnaire for use in epidemiologic studies. *Aging (Milano)*. 1995;7(5):358–366.
14. Von Korff M, Jensen MP, Karoly P. Assessing global pain severity by self-report in clinical and health services research. *Spine (Phila Pa 1976)*. 2000;25(24):3140–3151.
15. Ware Jr JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30(6):473–483.
16. Bot SD, Terwee CB, van der Windt DA, et al. Internal consistency and validity of a new physical workload questionnaire. *Occup Environ Med*. 2004;61(12):980–986.
17. Dallner M, Elo A-L, Gamberale F, et al. Validation of the general Nordic questionnaire (QPSNordic) for psychological and social factors at work. In: Nordic Council of Ministers, ed. *Vol Nord 200:12*. Copenhagen, Denmark: Nordic Council of Ministers; 2000.
18. Bouwmans C, Hakkaart-van Roijen L, Koopmanschap M, Krol M, Severens H, Brouwer W. Manual: iMTA Productivity Costs Questionnaire. www.imta.nl/questionnaires; 2013.
19. Koopmanschap MA. PRODISQ: a modular questionnaire on productivity and disease for economic evaluation studies. *Expert Rev Pharmacoecon Outcomes Res*. 2005;5(1):23–28.
20. van Roijen L, Essink-Bot ML, Koopmanschap MA, Bonsel G, Rutten FF. Labor and health status in economic evaluation of health care. The Health and Labor Questionnaire. *Int J Technol Assess Health Care*. 1996;12(3):405–415.
21. Bouwmans C, De Jong K, Timman R, et al. Feasibility, reliability and validity of a questionnaire on healthcare consumption and productivity loss in patients with a psychiatric disorder (TiC-P). *BMC Health Serv Res*. 2013;13:217.
22. Meerding WJ, IJzelenberg W, Koopmanschap MA, Severens JL, Burdorf A. Health problems lead to considerable productivity loss at work among workers with high physical load jobs. *J Clin Epidemiol*. 2005;58(5):517–523.
23. Kanters T. Questionnaires for the measurement of costs in economic evaluations, www.imta.nl/questionnaires. 2018.
24. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018;27(5):1159–1170.
25. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1171–1179.
26. Pallant J. *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS*. 6th ed. Sydney, Australia: Allen & Unwin; 2016.
27. Nunnally J, Bernstein I, eds. *Psychometric Theory*. 3rd ed. New York, NY: McGraw-Hill; 1994.
28. Kigozi J, Lewis M, Jowett S, Barton P, Coast J. Construct validity and responsiveness of the single-item presenteeism question in patients with lower back pain for the measurement of presenteeism. *Spine (Phila Pa 1976)*. 2014;39(5):409–416.
29. Lamers LM, Meerding WJ, Severens JL, Brouwer WB. The relationship between productivity and health-related quality of life: an empirical exploration in persons with low back pain. *Qual Life Res*. 2005;14(3):805–813.
30. Canjuga M, Hammig O, Bauer GF, Laubli T. Correlates of short- and long-term absence due to musculoskeletal disorders. *Occup Med (Lond)*. 2010;60(5):358–361.
31. IJzelenberg W, Molenaar D, Burdorf A. Different risk factors for musculoskeletal complaints and musculoskeletal sickness absence. *Scand J Work Environ Health*. 2004;30(1):56–63.
32. d'Errico A, Viotti S, Baratti A, et al. Low back pain and associated presenteeism among hospital nursing staff. *J Occup Health*. 2013;55(4):276–283.
33. Steenstra IA, Verbeek JH, Heymans MW, Bongers PM. Prognostic factors for duration of sick leave in patients sick listed with acute low back pain: a systematic review of the literature. *Occup Environ Med*. 2005;62(12):851–860.
34. Kresal F, Suklan J, Roblek V, Jerman A, Mesko M. Psychosocial risk factors for low back pain and absenteeism among Slovenian professional drivers. *Cent Eur J Public Health*. 2017;25(2):135–140.
35. Janssens H, Clays E, de Clercq B, et al. Association between psychosocial characteristics of work and presenteeism: a cross-sectional study. *Int J Occup Med Environ Health*. 2016;29(2):331–344.
36. Zhang W, Bansback N, Kopec J, Anis AH. Measuring time input loss among patients with rheumatoid arthritis: validity and reliability of the Valuation of Lost Productivity questionnaire. *J Occup Environ Med*. 2011;53(5):530–536.
37. Lardon A, Dubois JD, Cantin V, Piche M, Descarreaux M. Predictors of disability and absenteeism in workers with non-specific low back pain: a longitudinal 15-month study. *Appl Ergon*. 2018;68:176–185.
38. Andresen EM. Criteria for assessing the tools of disability outcomes research. *Arch Phys Med Rehabil*. 2000;81(12 suppl 2):S15–S20.
39. Altman DG. *Practical Statistics for Medical Research*. London, UK: Chapman and Hall; 1991.
40. Zhang W, Anis AH. Health-related productivity loss: NICE to recognize soon, good to discuss now. *Pharmacoeconomics*. 2014;32(5):425–427.
41. Braakman-Jansen LM, Taal E, Kuper IH, van de Laar MA. Productivity loss due to absenteeism and presenteeism by different instruments in patients with RA and subjects without. *RA. Rheumatology (Oxford)*. 2012;51(2):354–361.
42. Zhang W, Gignac MA, Beaton D, et al. Productivity loss due to presenteeism among patients with arthritis: estimates from 4 instruments. *J Rheumatol*. 2010;37(9):1805–1814.
43. Beaton DE, Tang K, Gignac MA, et al. Reliability, validity, and responsiveness of five at-work productivity measures in patients with rheumatoid arthritis or osteoarthritis. *Arthritis Care Res (Hoboken)*. 2010;62(1):28–37.

Paper IV

Killingmo RM, Tveter AT, Småstuen MC, Storheim K, Grotle M

Comparison of self-reported and public registered absenteeism among people on long-term sick leave due to musculoskeletal disorders: criterion validity of the iMTA Productivity Cost Questionnaire

Eur J Health Econ. 2021 Aug; 22(6):865-872. DOI: 10.1007/s10198-021-01294-0

The European Journal of Health Economics (2021) 22:865–872
https://doi.org/10.1007/s10198-021-01294-0

ORIGINAL PAPER

Comparison of self-reported and public registered absenteeism among people on long-term sick leave due to musculoskeletal disorders: criterion validity of the iMTA Productivity Cost Questionnaire

Rikke Munk Killingmo[1] · Anne Therese Tveter[1,2] · Milada C. Småstuen[1] · Kjersti Storheim[1,3] · Margreth Grotle[1,3]

Received: 8 July 2020 / Accepted: 16 March 2021 / Published online: 1 June 2021

Abstract

Objective To evaluate criterion validity of the iMTA Productivity Cost Questionnaire (iPCQ) by comparing iPCQ-reported occurrence and duration of long-term absenteeism (> 4 weeks) with public registry data collected from the Norwegian Labour and Welfare Administration (NAV) among people on sick leave due to musculoskeletal disorders.

Method Baseline data from a cohort study was used, in which people on sick leave for at least 4 weeks due to musculoskeletal disorders were recruited electronically through the NAV website. To compare the occurrence of long-term absenteeism overall agreement between the two methods was measured by percentages. To compare the duration (number of days with absenteeism) and adjusted duration (number of days with complete absenteeism) of long-term absenteeism we conducted intraclass correlation coefficient (ICC) two-way random average agreement, descriptive statistic and Wilcoxon signed-rank test.

Results In total, 144 participants with a median age (range) of 49 (24–67) were included. The overall agreement on the occurrence of long-term absenteeism was 100%. The ICC value was 0.97 and 0.86 for duration and adjusted duration of long-term absenteeism, respectively. The median difference$_{(\text{iPCQ-registry})}$ between the two methods was 0 and 17 days for long-term absenteeism duration and adjusted duration, respectively. A significant difference between the two methods was observed (Wilcoxon signed-rank test, $p < 0.001$) with regards to adjusted duration of long-term absenteeism.

Conclusion The iPCQ showed good agreement with public registry data regarding the occurrence and duration of long-term absenteeism among people with musculoskeletal disorders on long-term sick-leave in Norway. However, the iPCQ does not cover part-time sick-leave and thereby potentially overestimate the total amount of long-term absenteeism.

Trial registration ClinicalTrials.gov Identifier no. NCT04196634.

Keywords Productivity costs · Absenteeism · Measurement properties · Musculoskeletal disorders

JEL classification B41

Introduction

The impact of disease and disorder on productivity is an important part of health economic evaluations. When a societal perspective is included in research, it can provide information on the relative costs of different disorders and on the relative cost-effectiveness and/or cost-utility of health care interventions. Thus, valid information on productivity costs is crucial in health economic evaluations and decision-making on how to best allocate resources [1–4]. Currently, there is no gold standard for measuring

✉ Rikke Munk Killingmo
rikke@oslomet.no

1 Department of Physiotherapy, Oslo Metropolitan University, Oslo, Norway

2 National Advisory Unit on Rehabilitation in Rheumatology, Division of Rheumatology and Research, Diakonhjemmet Hospital, Oslo, Norway

3 Division of Clinical Neuroscience, Research and Communication Unit for Musculoskeletal Health (FORMI), Oslo University Hospital, Oslo, Norway

productivity costs [2, 5, 6]. Nonetheless, there is a general agreement that one should measure productivity costs related to both absence from paid work (absenteeism), reduced productivity while at paid work (presenteeism) and costs related to unpaid work, such as household work, care work and volunteer work [4].

The iMTA Productivity Cost Questionnaire (iPCQ) is a self-reported outcome measure recently developed to cover these three domains of productivity costs [4]. It was designed to capture core parts of the existing questionnaires and to be a short, generic outcome measure allowing for quantification and valuation of all productivity costs in a single instrument [4]. Three studies have tested some of the measurement properties of the iPCQ [4, 7, 8]. Bouwmans et al. [4] confirmed its feasibility and face validity. Munk et al. [8] investigated and demonstrated overall good content and construct validity and reliability. In a modified version (iPCQ-VR), Beemster et al. [7] tested reliability, agreement and responsiveness of the core parts of absenteeism and presenteeism; they found good measurement properties on long-term sick leave, and poor measurement properties on short-term sick leave and presenteeism. To the best of our knowledge, the original iPCQ version has not been tested with respect to criterion validity. Testing criterion validity of iPCQ self-reported long-term absenteeism is feasible by validating against public registry data, which might be considered as a "golden standard". Testing criterion validity of the remaining domains (presenteeism and costs related to unpaid work) poses significant challenges due to the lag of a "gold standard" or objective measures [9].

Therefore, the aim of this study was to evaluate criterion validity of the iPCQ by comparing self-reported occurrence and duration of long-term absenteeism, assessed with the Norwegian iPCQ [8], with public registry data collected among people on long-term sick leave due to musculoskeletal disorders. A population group we consider to be relevant for this study, as musculoskeletal disorders is one of the leading causes of disability worldwide [10] accounting for a huge amount of productivity costs [11]).

Method

Design and setting

The present study was part of a prospective observational cohort study among people on sick leave due to musculoskeletal disorders (the MI-NAV project), conducted within the Norwegian Labour and Welfare Administration (NAV) [12]. Baseline data from the cohort study was compared with public registry data with respect to occurrence and duration of long-term absenteeism.

Participants and recruitment procedure

Eligible participants were people on sick leave for at least 4 weeks due to musculoskeletal disorders, aged 18 or above. Exclusion criteria were people being unable to read or write in Norwegian or English and people on sick leave longer than a 12-month period retrospectively from baseline. Recruitment of participants and consenting to participation was performed electronically through a link on everyone's individual profile page at the NAV website. Recruitment was between November 2018 and Mars 2019.

The Mi-NAV project was classified as a quality assessment study by the Norwegian Regional Committee for Medical Research Ethics (Reference No. 2018/1326/REK sør-øst A) and approved by the Norwegian Centre for Research Data (NSD 861249) in 2018.

Measurements

At baseline, the included participants completed a comprehensive questionnaire covering sociodemographic variables (sex, age, education level and mother tongue) and pain intensity in addition to self-reported long-term absenteeism by the iPCQ [4]. The Numeric Rating Scale (NRS 0–10) was used to measure pain intensity [13]. In addition, public registry data on long-term absenteeism as well as the related diagnostic code was collected from the Norwegian Labour and Welfare Administration (NAV), in the period from baseline to 12 months retrospectively.

The iMTA Productivity Cost Questionnaire

The iPCQ consists of 18 items and adopts a recall period of 4 weeks (except for item no. 5 and 6). In the introduction, nine items assess the date of reply and the following sociodemographic factors: age, sex, education level, work status, paid or unpaid work, profession, number of workdays and work hours per week of paid work. Further, productivity costs are measured in three separate index scores with individual sum scores: absence from paid work (absenteeism, with a distinction between short- (≤ 4 weeks) and long-term (> 4 weeks) absenteeism), reduced productivity at paid work (presenteeism) and productivity loss in unpaid work [14]. The occurrence and duration of long-term absenteeism can be calculated from items no. 5 and 6 ("Did you miss work earlier than the period of 4 weeks due to being sick? This is referring to one whole uninterrupted period of missed work as a result of being sick." (no, yes). "If yes, when did you call in sick?" (day, month, year).

The Norwegian versions as well as the manual for the iPCQ are available from the Institute for Medical Technology Assessment (iMTA) at Erasmus University Rotterdam [15].

Registry data

NAV is the public welfare agency in Norway. Workers in Norway qualify for sickness benefits from NAV if they have been in paid work for the last 4 weeks before the sickness incident, and if the occupational disability is documented by a doctor's sick leave certificate. In general, sickness benefit (100% of salary) can be received from the first day of reported sick and up to 1 year. If the person is still unable to work after 1 year, he or she may be entitled to work assessment allowance or disability benefits.

The data on absenteeism collected from the NAV registry contains dates and grading of absenteeism as well as the diagnostic codes related to the absence.

Outcomes

The outcomes in the present study will be occurrence and duration of long-term absenteeism. The occurrence of long-term absenteeism is defined as whether a continuous period of more than 4 weeks of absenteeism is recorded retrospectively from baseline (yes/no). The duration of long-term absenteeism is defined as the duration of a continuous period of absenteeism from baseline to maximum 12 months retrospectively. The duration of long-term absenteeism will be operationalized in two different ways (1) by calculating number of calendar days from start date until end date of sick leave (defined as the date the iPCQ was completed) (duration) and (2) by adjusting for grading of absenteeism, summarizing number of days with part-time sick leave to number of days with complete sick leave (adjusted duration) (e.g., 10 days with 50% sick leave equals absenteeism duration and adjusted duration of 10 and 5 days, respectively).

Analyses

To assess criterion validity, the COSMIN group recommends evaluating the extent to which an instrument is an adequate reflection of a "gold standard" [16, 17]. To compare the occurrence of long-term absenteeism participants were classified according to whether a continuous period of long-term absenteeism had been recorded by the iPCQ (yes/no) and the registry (yes/no). The overall agreement between the two methods was expressed as follows: OA = (number of identical/total answers) × 100.

To compare the duration and adjusted duration of long-term absenteeism, we computed intraclass correlation coefficient (ICC) using two-way random average agreement. The acceptable level of ICC was set to > 0.70 [16]. In addition, to illustrate the relationship between the two methods, we depicted the differences$_{(\text{iPCQ-registry})}$ and averages of these using Blant–Altman plots. Also, the differences$_{(\text{iPCQ-registry})}$ were described with medians and interquartile ranges and analyzed with the Wilcoxon signed rank test. To test whether differences between the two methods were associated with the length of sick leave, as recorded in the registry, stratified analyses for the following categories of absenteeism length were performed: ≤ 3 months, >3 months to ≤ 6 months and ≥6 months. In addition, Spearman's rho was used to assess the correlation between the differences$_{(\text{iPCQ-registry})}$ and the length of sick leave. Correlation coefficients smaller than 0.3, between 0.3 and 0.6 and greater than 0.6 were considered low, moderate and high, respectively [18].

To test credibility of the primary analyses, sensitivity analyses without outliers were performed. Outliers were identified with simple scatter plots by visual inspection.

All data analyses were performed using SPSS version 24 (SPSS Inc., Chicago, IL, USA).

Results

A total of 144 participants with a median age (range) of 49 (24–67) had complete data for the current analyses and were included in this study. Almost half of the included participants had high education level and 59% were females. On average, they reported moderate pain, and their absenteeism was most frequently related to musculoskeletal disorders in the upper limbs. The study sample characteristics are shown in Table 1.

Self-reported occurrence of long-term absenteeism assessed with the iPCQ was identical to data retrieved from the registry; thus, the two methods revealed a 100% agreement.

Self-reported duration and adjusted duration of long-term absenteeism assessed with the iPCQ correlated highly and acceptably with data retrieved from the registry. The ICC (95%CI) were 0.93 (0.91–0.95) and 0.75 (0.48–0.86) for duration and adjusted duration of long-term absenteeism, respectively. A sensitivity analysis excluding 4 outliers confirmed these results with ICC (95% CI) values of 0.99 (0.99–0.99) and 0.83 (0.57–0.91) for duration and adjusted duration of long-term absenteeism, respectively. Figures 1 and 2 illustrate the differences$_{(\text{iPCQ-registry})}$ plotted against data from the registry including the 95% limits of agreement.

Descriptive statistics for the duration and adjusted duration of long-term absenteeism is presented in Table 2. With regards to the duration of long-term absenteeism there was a median difference$_{(\text{iPCQ-registry})}$ of 0 days and the two methods did not differ significantly (Wilcoxon signed-rank test, $p = 0.064$). A sensitivity analysis excluding the

Table 1 Participants demographic characteristics and clinical status ($n = 144$)

		Missing, *n* (%)
Female, *n* (%)	85 (59.0)	–
Age in years, median (range)	49 (24–67)	–
Education level high, *n* (%)	71 (49.3)	–
Mother tongue Norwegian, *n* (%)	128 (88.9)	–
Weekly workhours, median (IQR)	37.5 (25–37.5)	9 (6.3)
Weekly workdays, median (IQR)	5 (4–5)	9 (6.3)
Type of sick leave, *n* (%)	–	–
Partial sick leave	17 (11.8)	–
Complete sick leave	48 (33.3)	–
Partial and complete sick leave	79 (54.9)	–
Absenteeism longer than 4 weeks, *n* (%)	144 (100)	–
Presenteeism last 4 weeks, *n* (%)	68 (47.2)	1 (0.7)
Productivity loss unpaid work last 4 weeks, *n* (%)	75 (52.1)	1 (0.7)
Productivity cost (iPCQ index scores), median (IQR)	–	–
Absenteeism in hours	566 (380–894)	13 (9.0)
Presenteeism in hours	40 (16–72)	8 (11.8)
Productivity loss unpaid work in hours	28 (11–45)	6 (8.0)
Pain severity last 2 weeks (NRS 0–10), mean (SD)	5 (2)	–
Pain location, *n* (%)	–	–
Upper limbs	41 (28.5)	–
Lower limbs	22 (15.3)	–
Back and neck	28 (19.4)	–
Multiple pain areas	30 (20.8)	–
Others	23 (15.9)	

Pain location is based on diagnostic code related to absenteeism collected from the The Norwegian Labour and Welfare administration registry

iPCQ Institute for Medical Technology Assessment Productivity Cost Questionnaire, *IQR* interquartile range, *NRS* Numeric Rating Scale

4 outliers provided the same result (Wilcoxon signed-rank test, $p = 0.274$). With regards to the adjusted duration of long-term absenteeism the degree of agreement between the two methods was poorer (Table 2). When compared with the registry the participants overestimated the numbers of days with long-term absenteeism with median 17 days, and a statistically significant difference between the two methods was revealed (Wilcoxon signed-rank test, $p < 0.001$). A sensitivity analysis excluding the 4 outliers provided the same result (Wilcoxon signed-rank test, $p < 0.001$).

Descriptive statistics for the duration and adjusted duration of long-term absenteeism, categorized by the length of sick leave is presented in Table 3. With regards to the adjusted duration of long-term absenteeism descriptive statistic indicated that the $\text{difference}_{(\text{iPCQ-registry})}$ between the two methods increased with the length of sick leave. However, formal testing with the Spearman's rho only revealed a moderate correlation between the two variables (rho = 0.44).

Discussion

In this study, we found that self-reported occurrence and duration of long-term absenteeism assessed with the iPCQ was an adequate reflection of public register data. However, with regards to adjusted duration of long-term absenteeism the iPCQ overestimated the number of days with complete sick leave as compared to public registry data.

Our results regarding self-reported and registered occurrence of long-term absenteeism are in line with other studies. Grøvle et al. [19] showed an overall agreement of 85% between self-reported and registry data on occurrence of absenteeism among patients with sciatica. Likewise, in a cohort on employees in Swedish public sector, Voss et al. [20] reported an overall agreement of 74–91%.

Previous studies [7, 8] have illuminated that the iPCQ does not cover part-time sick leave and thereby potentially lead to an overestimation of the total amount of absenteeism, including related costs. Therefore, we decided to

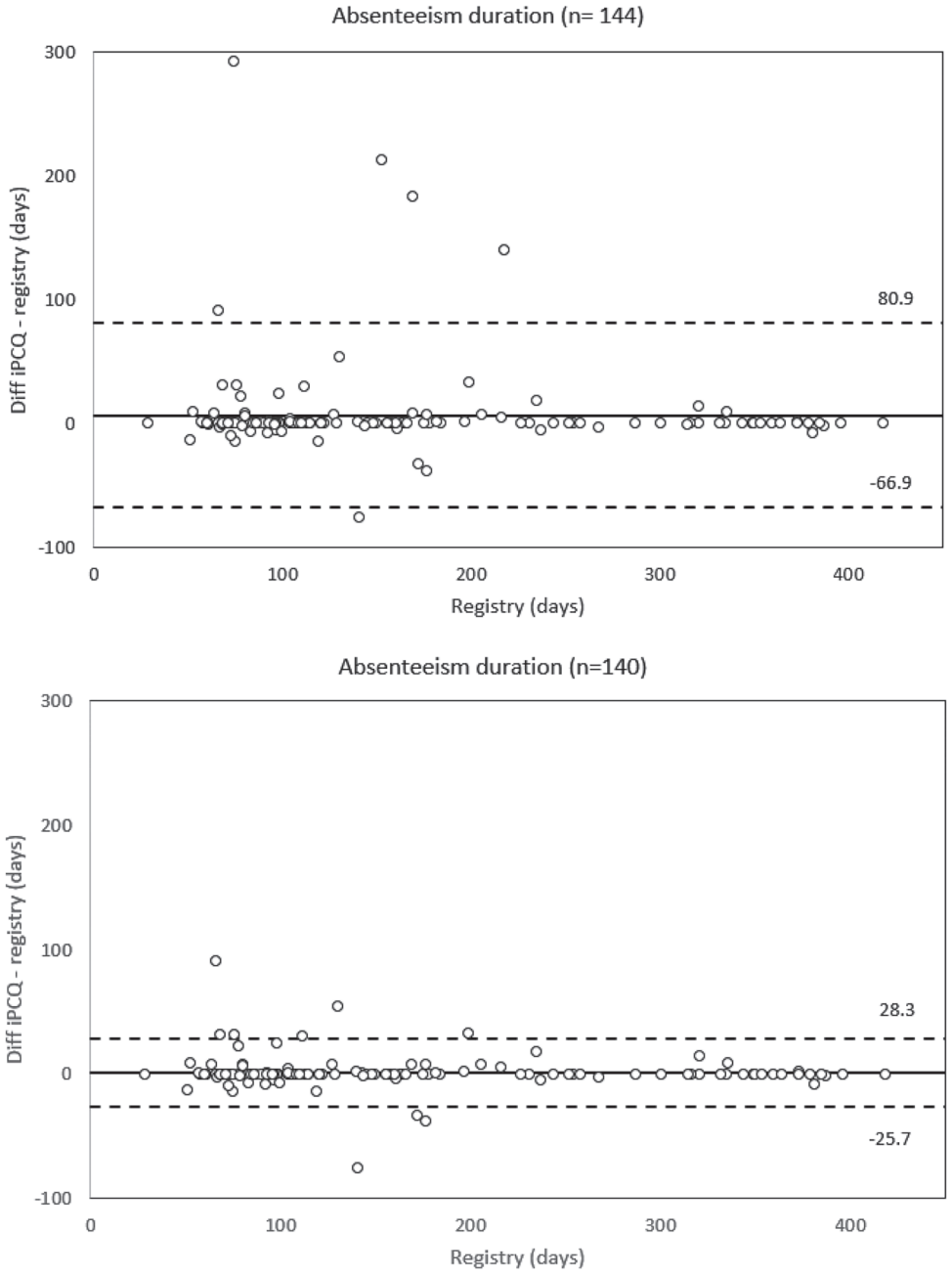

Fig. 1 The difference between iPCQ and registry-recorded long-term absenteeism duration plotted against the registry-recorded data. The central horizontal line represents the mean difference, the flanking lines represent the 95% limits of agreement

operationalize the duration of long-term absenteeism in two different ways (duration and adjusted duration). With regards to duration of long-term absenteeism our results are in line with other studies. A recent meta-analysis supports a satisfactory agreement between self-reported and registry data on duration of absenteeism, though people in most studies have a tendency of underreporting [3]. To the best of our knowledge, our study is the first to compare self-reported and registered adjusted duration of long-term absenteeism. However, it seems reasonable that a measuring toll not covering part-time sick leave tends to overestimate the total amount and of long-term absenteeism, including related costs. Furthermore, that longer time periods of absenteeism lead to larger differences.

The main limitation of this study is that we did not evaluate criterion validity of short-term absenteeism. However, it is likely to assume that short-term absenteeism is less biased, as shown previously [19]. A second potential weakness of this study is the lack of information regarding accuracy of the NAV registry. Because criterion validity is concerned with how well an instrument is an adequate reflection of a "gold standard" [16, 17] it is questionable to what degree the NAV registry can be used to provide evidence for criterion validity. However, because it composes the basis

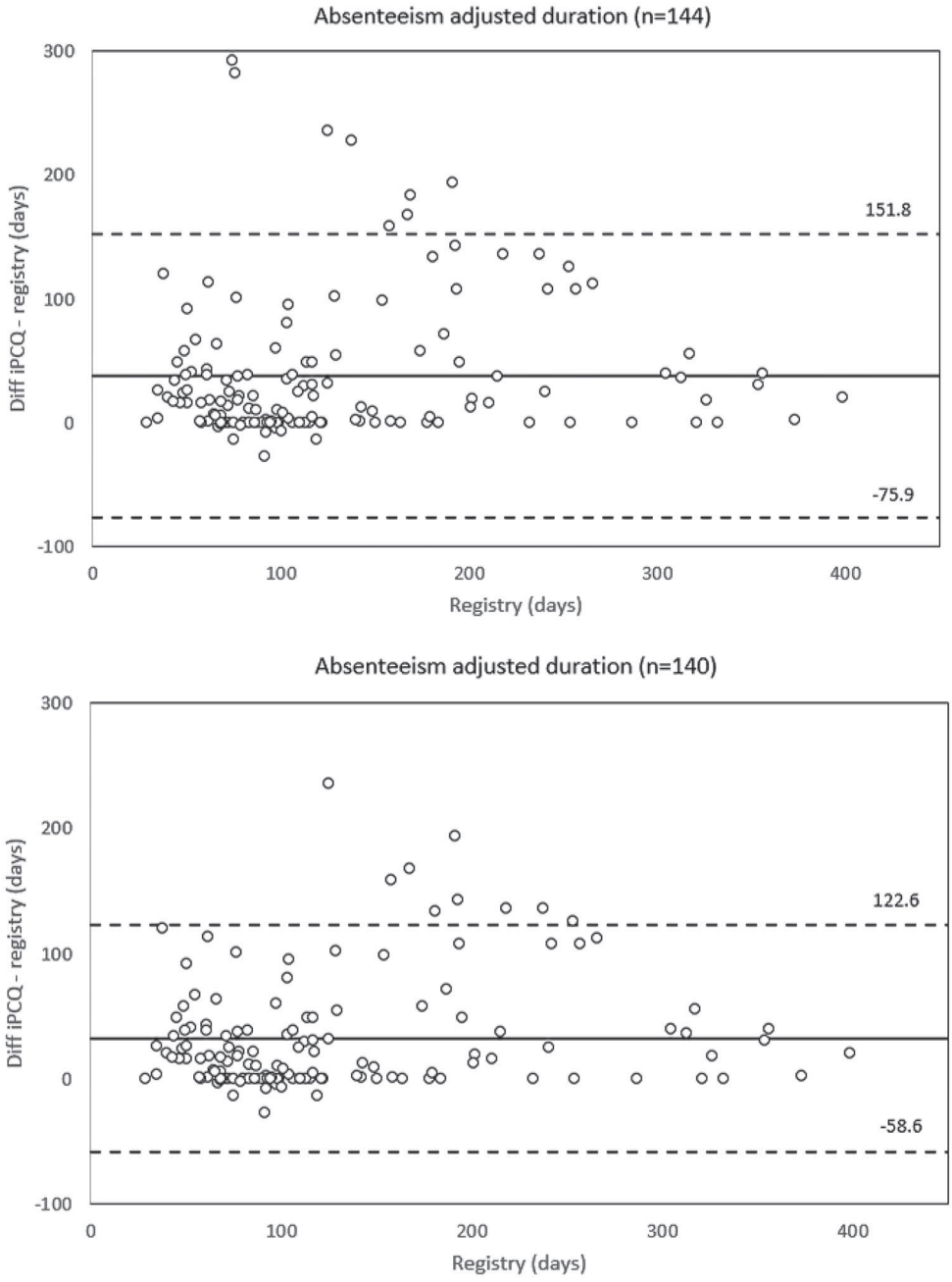

Fig. 2 The difference between iPCQ and registry-recorded long-term absenteeism adjusted duration plotted against the registry-recorded data. The central horizontal line represents the mean difference, the flanking lines represent the 95% limits of agreement

for payment of sickness benefits in Norway, it is generally regarded as accurate. A third weakness is the lack of data on eligible participants choosing not to participate. Owing to limited resources, it was not possible to record information on all eligible participants during the data collection period. However, this comparison will be carried out at a later stage in the MI-NAV project.

The main strength of the present study is that it is the first to test criterion validity of self-reported long-term absenteeism with the iPCQ and that this validation was conducted in line with COSMIN guidelines [17]. Furthermore, that we evaluated the implication of part-time sick leave.

Conclusion

In conclusion, this study showed that self-reported occurrence and duration of long-term absenteeism assessed with the iPCQ have good agreement with public registry data collected from the NAV among people on long-term sick leave due to musculoskeletal disorders in Norway. Nevertheless, the iPCQ does not cover part-time sick-leave and thereby potentially overestimates the total value of productivity costs related to long-term absenteeism. Since the

Table 2 Parameters of long-term absenteeism duration and adjusted duration

	Main sample ($n=144$)	Sensitivity analysis ($n=140$)
Absenteeism duration (days), median (IQR)		
iPCQ	121 (85–232)	115 (84–219)
Registry	115 (80–204)	114 (80–204)
Difference, iPCQ – registry	0 (0–0)	0 (0–0)
Absenteeism adjusted duration (days), median (IQR)		
iPCQ	121 (85–232)	115 (84–219)
Registry	100 (71–167)	100 (70–167)
Difference, iPCQ – registry	17 (0–49)	16 (0–41)

Absenteeism duration is calculated by subtracting the start date from the end date of sick leave. Absenteeism adjusted duration is calculated by adjusting for partial sick leave, summarizing number of days with partial sick leave to number of days with complete sick leave

iPCQ Institute for Medical Technology Assessment Productivity Cost Questionnaire, *IQR* interquartile range

Table 3 Parameters of long-term absenteeism duration and adjusted duration for absenteeism periods of different lengths

	≤ 3-Mo ($n=44$)	> 3-Mo to ≤ 6-Mo ($n=60$)	> 6-Mo ($n=40$)
Absenteeism duration (days), median (IQR)			
iPCQ	72 (65–85)	121 (98–155)	327 (246–363)
Registry	72 (66–80)	120 (99–155)	319 (236–363)
Difference, iPCQ – registry	0 (0–1)	0 (0–0)	0 (0–1)
Absenteeism adjusted duration (days), median (IQR)			
iPCQ	72 (65–85)	121 (98–155)	327 (246–363)
Registry	68 (50–74)	104 (86–121)	217 (181–300)
Difference, iPCQ – registry	2 (0–20)	10 (0–39)	57 (19–122)

Absenteeism duration is calculated by subtracting the start date from the end date of sick leave. Absenteeism adjusted duration is calculated by adjusting for partial sick leave, summarizing number of days with partial sick leave to number of days with complete sick leave

iPCQ Institute for Medical Technology Assessment Productivity Cost Questionnaire, *IQR* interquartile range, *Mo* month

iPCQ is a generic instrument also measuring short-term absenteeism, further studies should validate it in other populations and among people on short-term sick leave.

Acknowledgements We would like to thank the Norwegian Labour and Welfare Administration for the possibility to include participants through their system.

Funding Open access funding provided by OsloMet - Oslo Metropolitan University. This study supported by the Research Council of Norway (the MI-NAV project), through the program "Sickness absence, work and health" and by Oslo Metropolitan University.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Burdorf, A.: Economic evaluation in occupational health—its goals, challenges, and opportunities. Scand. J. Work Environ. Health **33**(3), 161–164 (2007)
2. Krol, M., Brouwer, W., Rutten, F.: Productivity costs in economic evaluations: past, present, future. Pharmacoeconomics **31**(7), 537–549 (2013)
3. Johns, G., Miraglia, M.: The reliability, validity, and accuracy of self-reported absenteeism from work: a meta-analysis. J. Occup. Health Psychol. **20**(1), 1–14 (2015)
4. Bouwmans, C., et al.: The iMTA Productivity Cost Questionnaire: a standardized instrument for measuring and valuing health-related productivity losses. Value Health **18**(6), 753–758 (2015)
5. Tang, K.: Estimating productivity costs in health economic evaluations: a review of instruments and psychometric evidence. Pharmacoeconomics **33**(1), 31–48 (2015)
6. Zhang, W., Bansback, N., Anis, A.H.: Measuring and valuing productivity loss due to poor health: a critical review. Soc. Sci. Med. **72**(2), 185–192 (2011)
7. Beemster, T.T., et al.: Test–retest reliability, agreement and responsiveness of productivity loss (iPCQ-VR) and healthcare utilization (TiCP-VR) questionnaires for sick workers with chronic musculoskeletal pain. J. Occup. Rehabil. **29**, 91–103 (2019). https://doi.org/10.1007/s10926-018-9767-9
8. Munk, R., et al.: Measuring productivity costs in patients with musculoskeletal disorders: measurement properties of the Institute for Medical Technology Assessment Productivity Cost Questionnaire. Value Health **22**(12), 1410–1416 (2019)
9. Zhang, W., Anis, A.H.: Health-related productivity loss: NICE to recognize soon, good to discuss now. Pharmacoeconomics **32**(5), 425–427 (2014)
10. Global Burden of Disease Study, C: Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet **386**(9995), 743–800 (2015)
11. Bevan, S.: Economic impact of musculoskeletal disorders (MSDs) on work in Europe. Best Pract. Res. Clin. Rheumatol. **29**(3), 356–373 (2015)
12. Tveter, A.T., et al.: Risk assessment for prolonged sickness absence due to musculoskeletal disorders: protocol for a prospective cohort study. BMC Musculoskelet. Disord. **21**(1), 326 (2020)
13. Von Korff, M., Jensen, M.P., Karoly, P.: Assessing global pain severity by self-report in clinical and health services research. Spine (Phila Pa 1976) **25**(24), 3140–3151 (2000)
14. Bouwmans, C., et al.: Manual: iMTA Productivity Costs Questionnaire. 2013, Institute for Medical Technology Assessment. http://www.imta.nl/questionnaires
15. Kanters, T.: Questionnaires for the measurement of costs in economic evaluations. 2019 [cited 2019 09.10.2019]
16. Terwee, C.B., et al.: Quality criteria were proposed for measurement properties of health status questionnaires. J. Clin. Epidemiol. **60**(1), 34–42 (2007)
17. Mokkink, L.B., et al.: COSMIN study design checklist for patient-reported outcome measurement instruments. 2019: http://www.cosmin.nl
18. Andresen, E.M.: Criteria for assessing the tools of disability outcomes research. Arch. Phys. Med. Rehabil. **81**(12 Suppl 2), S15-20 (2000)
19. Grovle, L., et al.: Poor agreement found between self-report and a public registry on duration of sickness absence. J. Clin. Epidemiol. **65**(2), 212–218 (2012)
20. Voss, M., et al.: Comparisons of self-reported and register data on sickness absence among public employees in Sweden. Occup. Environ. Med. **65**(1), 61–67 (2008)

Errata list

Name of candidate Rikke Munk Killingmo

Title of thesis Cost of Illness due to Back Pain in Older People. Healthcare utilization, modifiable prognostic factors, and the measurement properties of self-reported productivity loss using the iMTA Productivity Cost Questionnaire

Page unnumbered	Para	Line	Original text	Correction	Corrected text
Summary page 1	3	5	aged >55 years	Cor	aged > 55 years
Sammendrag page 1	3	3	alderen ≥55 år	Cor	alderen ≥ 55 år
Sammendrag page 1	3	5	alderen >55 år	Cor	alderen > 55 år
Sammendrag page 2	2	5	Paracetamol	Cor	paracetamol
Sammendrag page 2	3	3	Paracetamol	Cor	paracetamol

Page	Para	Line	Original text	Correction	Corrected text
1	-	17	3.2 Protocol and registration …23	Cor	3.2 Protocol and registration …22
1	-	20	3.5 Eligibility criteria and recruitment procedure …24	Cor	3.5 Eligibility criteria and recruitment procedure …23
10	1	3	diseases [63]	Cor	diseases [63].
20	2	8	older back pain patients	Cor	older patients with back pain
23	1	7	to 12 months, retrospectively	Cor	to 12 months retrospectively
29	1	23	questionnaire CES-D [147].	Cor	questionnaire (CES-D) [147].
30	1	26	complaints [157, 158].	Cor	complaints [157, 158]
41	1	3-5	were performed on complete case data (in BACE-N). To assess credibility of the primary analyses in Paper II, sensitivity analyses without outliers were performed. Outliers were identified with simple scatter plots by visual inspection.	Cor	were performed on complete case data (in BACE-N).
43	-	21	(NRS 0-10)	Cor	(NRS, 0-10)
43	-	32	(CES-D 0-60)	Cor	(CES-D, 0-60)
43	-	33	(FABQ-PA 0-24)	Cor	(FABQ-PA, 0-24)
43	-	57	article for Paper III presented	Cor	article for Paper I presented
44	-	18	Pain severity last 1 (Study I) or 2 weeks (Study II) (NRS 0-10)	Cor	Pain severity last 1 (Paper III) or 2 weeks (Paper IV) (NRS, 0-10)
44	-	25	(SF36 0-100)	Cor	(SF36, 0-100)
44	-	28	(PWQ 0-100)	Cor	(PWQ, 0-100)
44	-	39	paper for Study I presented	Cor	article for Paper III presented
52	2	5	REMARk	Cor	REMARK
53	1	1	updated [5]s	Cor	updated [5]

Cor indicates correction of language