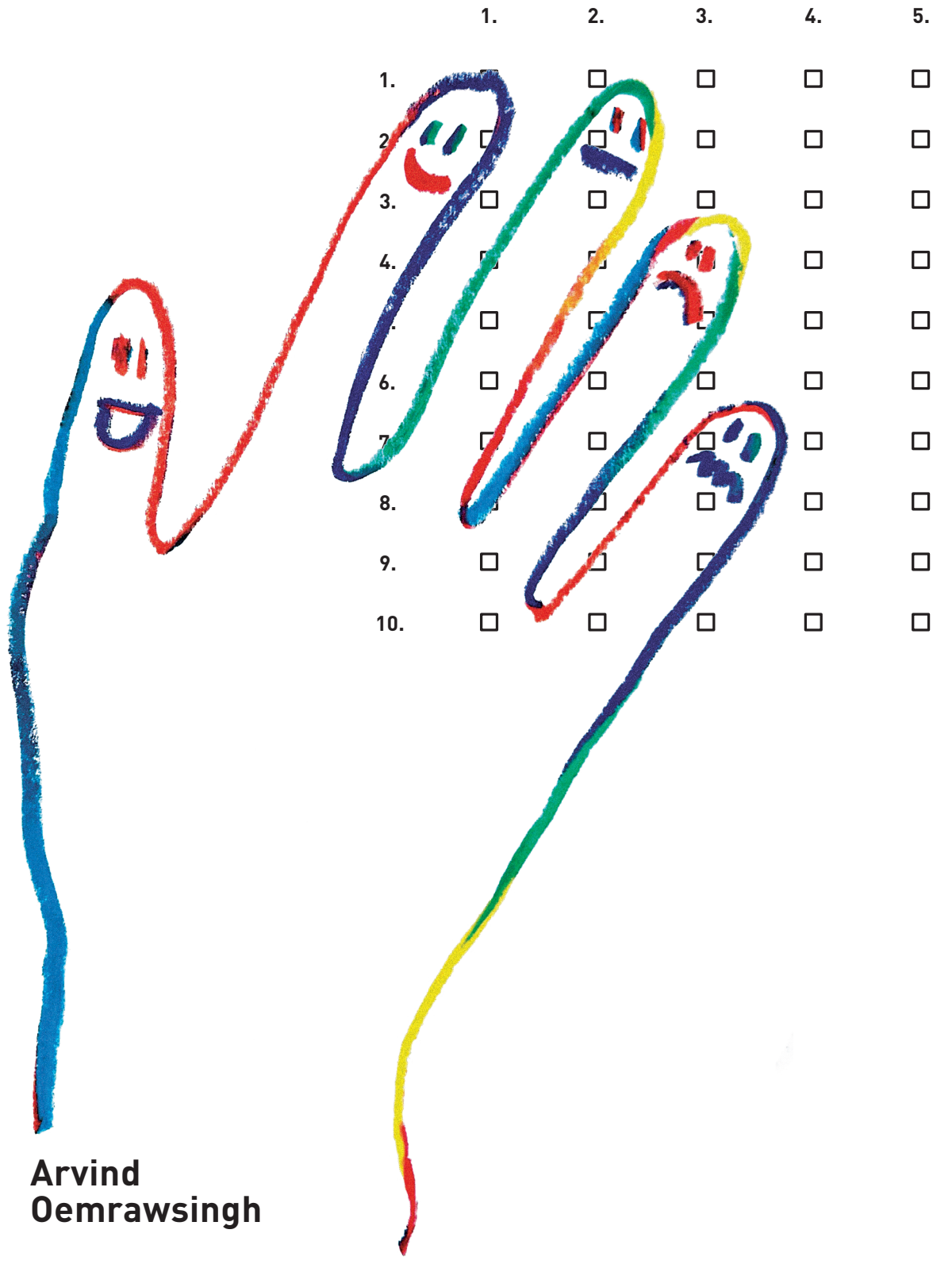
Patient-Reported Outcome Measures in Clinical Care: methodological aspects and practical applications



Arvind Oemrawsingh

Patient-Reported Outcome Measures in Clinical Care: Methodological Aspects and Practical Applications

by

Arvind Oemrawsingh

Rotterdam, the Netherlands
2021

ISBN: 978-94-6361-593-8

Patient-Reported Outcome Measures in Clinical Care: Methodological Aspects and Practical Applications

This thesis is partly realized thanks to the financial support of the Dutch Federation of University Medical Centers (NFU), Erasmus University Rotterdam, the Department of Surgery, and the Department of Public Health, Erasmus MC.

Layout & Print by Optima Grafische Communicatie in Rotterdam, the Netherlands (www.ogc.nl)
Cover design by Shortnotice Studio in Berlin, Germany (www.shortnotice.studio)

Patient-Reported Outcome Measures in Clinical Care: methodological aspects and practical applications

Patiënt-gerapporteerde uitkomstmaten in de klinische zorg: methodologische aspecten en praktische toepassingen

Proefschrift

Ter verkrijging van de graad van doctor aan de Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof. dr. A.L. Bredenoord
en volgens besluit van het College van Promoties.

De openbare verdediging zal plaatsvinden op
1 oktober 2021 om 13:00u.

door

Arvind Oemrawsingh
geboren te Plantation, Florida, USA

Erasmus University Rotterdam

DOCTORAL COMMITTEE

Promotors:	Prof. dr. J.A. Hazelzet
	Prof. dr. N.S. Klazinga
Other members:	Prof. dr. S. Siesling
	Prof. dr. J.J.M. Takkenberg
	Dr. P.J. Nederkoorn
Co-promotors:	Dr. N. van Leeuwen
	Dr. L.B. Koppert

Voor mijn ouders, Sunil & Maureen,
die dapper genoeg waren om mij “los te laten”.

“Wherever the art of Medicine is loved, there is also a love of Humanity.”
Hippocrates

LIST OF SCIENTIFIC PUBLICATIONS

This Thesis:

- van Egdom LSE[†], **Oemrawsingh A**[†], Verweij LM, Lingsma HF, Koppert LB, Verhoef C, Klazinga NS, Hazelzet JA. Implementing Patient-Reported Outcome Measures in Clinical Breast Cancer Care: A Systematic Review. *Value Health. 2019 Oct;22(10):1197-1226. doi: 10.1016/j.jval.2019.04.1927.*

- **Oemrawsingh A**, van Leeuwen N, Venema E, Limburg M, de Leeuw FE, Wijffels MP, de Groot AJ, Hilkens PHE, Hazelzet JA, Dippel DWJ, Bakker CH, Voogdt-Pruis HR, Lingsma HF. Value-Based Healthcare in Ischemic Stroke Care: Case-Mix Adjustment Models for Clinical and Patient-Reported Outcomes. *BMC Med Res Methodol. 2019 Dec 5;19(1):229. doi: 10.1186/s12874-019-0864-z.*

- **Oemrawsingh A**, Swami N, Valderas JM, Hazelzet JA, Pusic AL, Gliklich RE, Bergmark RW. Patient-Reported Morbidity Instruments: A Systematic Review. *Value Health. 2020 Jun;23(6):791-811. doi: 10.1016/j.jval.2020.02.006.*

- **Oemrawsingh A**, van Leeuwen N, van der Lugt A, Klazinga NS, Hazelzet JA, van Busschbach JJ, Dippel DWJ, Lingsma HF; on behalf of MR CLEAN investigators. Case-mix models for hospital comparisons in EQ-5D domains after ischemic stroke: a substudy of the MR CLEAN trial. *Submitted.*

- **Oemrawsingh A**, van Leeuwen N, Mureau MAM, Lingsma HF, Verhoef C, Klazinga NS, Hazelzet JA, Koppert LB. A Prospective Analysis of Short-Term Patient-Reported Outcomes After Breast-Conserving Therapy and Mastectomy Without Reconstruction. *Submitted.*

- **Oemrawsingh A**[†], Clarijs ME[†], Pusic AL, Lingsma HF, Verhoef C, Hazelzet JA, Koppert LB. BREAST-Q Breast-Conserving Therapy Module: Normative Data From A Dutch Sample of 9059 Women. *Accepted for publication.*

- Amini M[†], **Oemrawsingh A**[†], Verweij LM, Klazinga NS, Lingsma HF, Hazelzet JA, Eijkenaar F, van Leeuwen N. Facilitators and Barriers of Implementing Patient-Reported Outcome Measures in Clinical Care: an Academic Center's Initial Experience. *Health Policy. 2021 Jul 8:S0168-8510(21)00173-1. doi: 10.1016/j.healthpol.2021.07.001.*

Other Publications:

- **Oemrawsingh A**, de Boer NL, Brandt-Kerkhof ARM, Verhoef C, Burger JWA, Madsen EVE. Short-Term Complications in Elderly Patients Undergoing CRS and HIPEC: A Single Center's Initial Experience. *Eur J Surg Oncol. 2019 Mar;45(3):383-388. doi: 10.1016/j.ejso.2018.10.545.*

- van Kooten JP[ǂ], **Oemrawsingh A**[ǂ], de Boer NL, Verhoef C, Burger JWA, Madsen EVE, Brandt-Kerkhof ARM. Predictive Ability of C-Reactive Protein in Detecting Short-Term Complications After Cytoreductive Surgery and Hyperthermic Intraperitoneal Chemotherapy: A Retrospective Cross-Sectional Study. *Ann Surg Oncol. 2020 Jun 10. doi: 10.1245/s10434-020-08619-y.*

- **Oemrawsingh A**, Hazelzet JA, Koppert LB. "Hoofdstuk 17 Patiëntgerichte borstkankerzorg: ervaring en perspectief van een academisch centrum". NFU bundel *Gepersonaliseerde medische zorg – Innovatieve zorg afgestemd op persoonlijke behoeften, voorkeuren en waarden*. Juni, 2020. ISBN: 978-90-9033-183-6.

- Devarakonda KS, Timman R, Bouvy PF, **Oemrawsingh A**, Apon I, Mureau MAM, Koppert LB, Kranenburg LW. Trends in Emotional Functioning and Psychosocial Wellbeing in Breast Cancer Survivors. *Submitted.*

- Clarijs ME, **Oemrawsingh A**, Bröker MEE, Verhoef C, Lingsma HF, Koppert LB. Quality of Life of Caregivers of Breast Cancer Patients: a Cross-Sectional Evaluation. *Submitted.*

[ǂ] **= shared co-first authorship**

List of Commonly Used Abbreviations

VBHC	=	Value-Based Healthcare
PROM(s)	=	Patient-Reported Outcome Measure(s)
HRQoL	=	Health-Related Quality of Life
mRS	=	modified Rankin Scale
NIHSS	=	Netherlands Institute of Health Stroke Scale
EQ-5D	=	EuroQol-5D
EORTC-QLQ-C30	=	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire–Core 30
EORTC-QLQ-BR23	=	European Organization for Research and Treatment of Cancer Quality of Life Questionnaire–Breast Cancer 23
MR CLEAN	=	Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands

TABLE OF CONTENTS

Appendices

Chapter 1 General Introduction

VALUE-BASED HEALTHCARE

In many Western countries, there has been a rapidly increasing shift over the past decade of the current volume-driven healthcare system, fragmented and focused on the fee-for-service of provided healthcare services, towards "Value-Based Healthcare" (VBHC).[1,2] The goal of this patient-centered healthcare delivery system is to acquire high value – the ratio of patient health outcomes achieved to the costs of a full care cycle – for a specific medical condition.[1] In order to maximize the value for patients the current system, which is organized around what healthcare providers do, should shift its focus to what patients need and value.[3] In 2010, Michael Porter wrote in the *New England Journal of Medicine*: "Value should always be defined around the customer, and in a well-functioning health care system, the creation of value for patients should determine the rewards for all other actors in the system. Since value depends on results, not inputs, value in health care is measured by the outcomes achieved, not the volume of services delivered, and shifting focus from volume to value is a central challenge.[4]"

The (systematic) measurement of achieved health outcomes and costs per patient is one of six key inter-dependent components of value-based healthcare (Figure 1). These outcomes should matter to patients, should be measured for each medical condition and should cover the entire care cycle till completion.[3] Measurement of outcomes is central to improving the quality of care, and incentivizing care providers to improve the care.[5] In the Netherlands, institutions such as the Dutch National Health Care Institute (ZIN) and the Dutch Federation of University Medical Centers (NFU) have advocated the use of patient-reported data as feedback for improving health care services, promoting shared decision-making and delivering individualized care.[6,7]

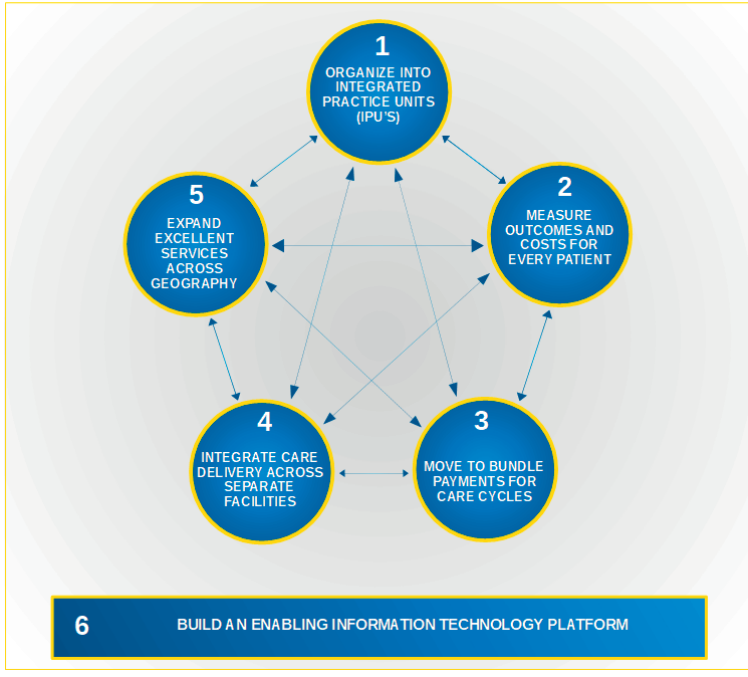


Figure 1 | Key Components of VBHC. Source: Porter, ME et al.[3], *Harvard Business Review*, 2013.

Porter proposed that a full outcome set for any medical condition can be categorized in a three-tiered hierarchy[4] (Figure 2):

- *Tier 1 Health status achieved or retained*: This tier represents outcomes related to survival and the degree of functional status achieved.
- *Tier 2 Process of recovery*: This tier represents outcomes related to the nature of the care cycle including the time to care completion/ recovery (e.g. rate of readmissions or emergency department visits) and return to normal daily activities, and the disutility of care.
- *Tier 3 Sustainability of health*: This tier represents outcomes related to the (post-recovery) sustainability of health, uninterrupted by recurrences or newly developed medical conditions.

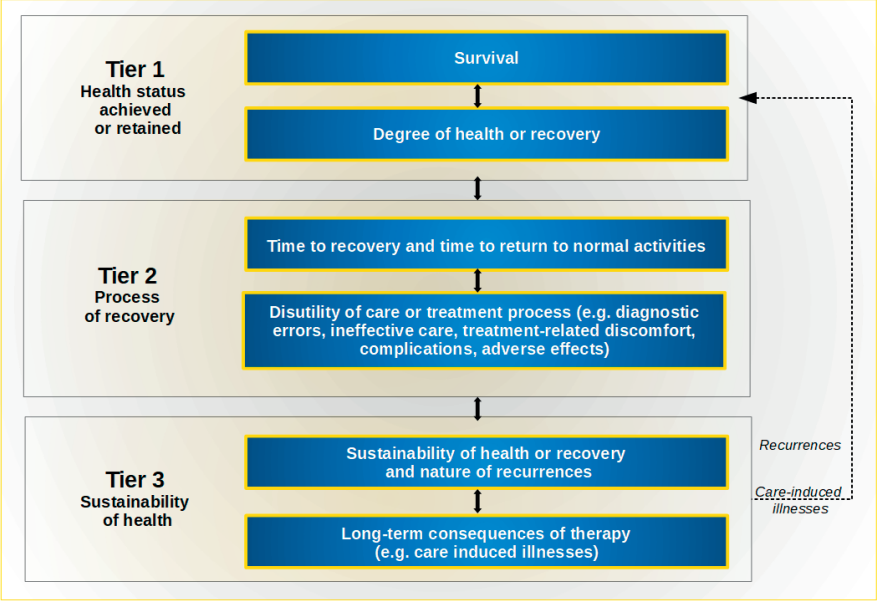


Figure 2 | Three-Tier Outcome Measures Hierarchy. Source: Porter, ME[4], *NEJM*, 2010.

Each tier contains two broad *levels*. Each of these levels has to be measured with at least one outcome metric. Levels can be discerned further into one or multiple *dimensions*. These (outcome) dimensions ultimately capture different aspects of patient health. It is important that appropriate time points and frequencies, with which to measure each outcome, are decided upon.

The utilization of measurements to assess quality of care is not a new concept.[8] In the 1850's, Florence Nightingale clearly provided statistical evidence that improving sanitation in hospitals, promoting hygiene standards, and preventing patient overcrowding led to a reduced mortality and morbidity in soldiers treated during the Crimean war.[9] Her visualized statistics convinced the British policymakers at that time that action was needed. The founder of evidence-based quality improvement is widely considered to be Ernest Codman, a surgeon from Massachusetts General Hospital and Harvard Medical School. He believed that prolonged patient observation and comprehensive measurement of patient outcomes following treatment was necessary to improve the care of future patients.[10] After initially being criticized by his colleagues, he was able to introduce standards for surgical practice within the College of Surgeons that formed the basis of today's health care accreditation programs by the Joint Commission International (JCI). Medical measures were developed in the 1950s in order to enable systematic hospital qual-

ity assessment.[11] In the 1960s, Avedis Donabedian proposed a conceptual model[12] comprising of:

- Structure measures e.g. ratio of physicians to patients
- Process measures e.g. percentage of women receiving mammograms
- Outcomes e.g. mortality

This aforementioned model has since become the paradigm for the modern era of quality measurement.[13] Up till now, quality measurement has mainly focused on clinical outcome indicators such as mortality and complications.

PATIENT-REPORTED OUTCOME MEASURES (PROMS)

Since the early 1990s the development (and consequent validation) of patient-reported measures, that captured the patients' perspective on their own health status, has increased exponentially.[14] These measures were initially developed for use in pharmaceutical and health services research and were largely restricted to England, Sweden, and parts of the US as a way to improve the clinical care of patients.[15-17] Through the years, these particular measures have gained an established term namely Patient-Reported Outcome Measures (PROMs).

Patient-Reported Outcomes (PROs) can be defined as direct feedback from the patient on his or her own health status (including health-related quality of life) and/or treatment, without external interpretation of patients' responses by healthcare professionals.[18] PROMs are the (standardized) instruments or tools or questionnaires or surveys used to capture PROs, which can encompass a myriad of areas such as symptoms, functional status, (health-related) quality of life, treatment effects, and (treatment) satisfaction.[17,19]

Patient-Reported Outcome Measures can be administered through a variety of methods: paper-and-pencil, interviews, or increasingly more frequent, electronic devices (e.g. (tablet) computer, mobile phones).[20] There is a plethora of various PROMs, which differ in the amount of questions ("items") asked, the nature and wording of questions asked, and the final scoring of questionnaires. The quality between instruments can vary substantially, which is usually attributed to differences in multiple measurement properties[21-23], such as:

- *Validity:* The extent to which the instrument's scores correlate with a "gold standard" measure (criterion validity), or the extent to which the instrument accurately measures the phenomenon (construct) it is intended to measure

(construct validity), or the extent to which items are representative of the content of the domain that the instrument is designed to measure (content validity).

- *Reliability:* The degree to which an instrument yields consistent results, regardless of occasion and/or setting.
- *Sensitivity/ Responsiveness:* The ability of the instrument to detect a change over time in the measured construct.
- *Interpretability:* The assignment of qualitative meaning to an instrument's score. In contrast to the aforementioned properties, interpretability is not considered a psychometric property.

PROMs can, broadly speaking, be classified as either generic, disease-specific ("condition-specific") or domain-specific instruments[22,24] (Figure 3):

- *Generic*: These measures capture health aspects that can pertain to many disease populations, which allow comparisons across various diseases, between and within interventions, and delivery of health services. Another advantage is that some generic PROMs are useful to evaluate the outcomes of health interventions in health economic studies (e.g. EQ-5D[25]), and to compare the performance of health service providers. Examples of widely used generic and multi-dimensional PROMs include the EQ-5D, SF-36 (Health Survey Short Form – 36)[26] and PROMIS (Patient-Reported Outcomes Measurement Information System)[27].

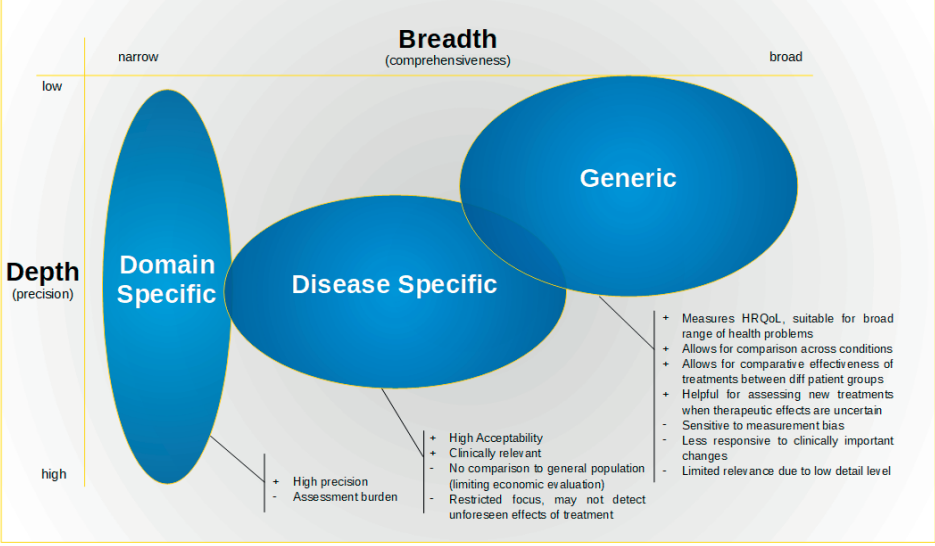


Figure 3 | Types of PROMs Including Advantages and Disadvantages. Source: Adapted from ICHOM.
Note. HRQoL = health-related quality of life

- *Disease-specific/ condition-specific*: These measures are intended to capture aspects that are important for one (or a few) particular disease(s) (e.g. stroke), set of diseases (e.g. cancer), a domain (e.g. pain) or an intervention (e.g. hip arthroplasty). These PROMs provide a more detailed insight into patients' self-reported health status, which make them more likely to be relevant in clinical care. However, a disadvantage of these measures is that they do not allow comparisons of health across patients with a variety of medical conditions.[22] Examples of disease-specific PROMs are the SS-QOL (Stroke-Specific Quality of Life)[28] and EORTC QLQ-C30 (Treatment of Cancer Quality-of-Life Questionnaire)[29].
- *Domain-specific*: These measures evaluate specific quality-of-life domains of a disease, e.g. pain, depression or fatigue.[30] These domains do not provide disease-specific information. An advantage of domain-specific PROMs is that they can be utilized to measure specific aspects of health across a broad array of diseases/ conditions. Domain-specific PROMs can measure a single domain (unidimensional) or multiple domains (multi-dimensional). Examples of domain-specific PROMs are the VAS (Visual Analogue Scale)[31], HADS (Hospital Anxiety and Depression Scale)[32] and FSS (Fatigue Severity Scale)[33].

Currently, both generic and disease-specific measures are considered as complementary and it has been recommended that both types of PROMs be collected, regardless of application.[21,22]

Patient-Reported Outcome Measures have increasingly been used over the years with various goals of screening for undetected symptoms/ issues[34], monitoring response to treatment and disease progression, improving patient outcomes[35], promoting individualized care[36], facilitating patient-provider communication and shared decision-making[37-40]. In addition, aggregated PROM data across patient populations, treatments, care providers or hospitals may be used as quality indicators for quality of care improvement and for transparency of outcomes across care providers through benchmarking, as part of promoting value in health care.[15,41,42] The interest in PROMs is reflected by the number of recently published reviews on the clinical application and impact of PROMs on various patient, process and health utilization outcomes.[43-46] Although the growing use of PROMs across different conditions in clinical care is encouraging[47,48], there are still critical questions that remain with regards to the collection, analysis, and interpretation of PROMs for quality of care improvement[41,49]. In addition, the relatively slow incorporation of PROMs in routine clinical care highlights the knowledge gaps and the complex-

ity of its implementation (development, introduction, testing, integration and monitoring of PROMs).

METHODOLOGICAL CHALLENGES

It is important to note that patient outcomes are not only a function of the quality of care, but also on patient characteristics that may influence response to care (e.g. risk factors/ morbidity), the natural course of a disease, and random chance.[50] Two important methodological issues that apply to patient-reported outcome measures, similar to other quality-of-care measures, are statistical uncertainty and the influence of case-mix (Figure 4). Statistical uncertainty relates to the variation in outcome between groups (e.g. treatments, hospitals, patient populations) caused just by random chance. Statistical uncertainty is dependent on the number of patients treated and the number of events (e.g. per treatment strategy or hospital), with smaller numbers of patients or events leading to larger amounts of statistical uncertainty. If statistical uncertainty is not taken into account during analysis, it may lead to the overestimation of outcome differences.[51]

Case-mix factors can be defined as patient characteristics, thus outside the influence of actions by the care provider, which may have an effect on (quality of care) outcomes. These case-mix factors may have differing distributions across groups, and should be taken into consideration when reporting on outcome differences.[52]

Figure 4 | Sources of Observed Between-Hospital Outcome Differences. Source: Lingsma HF[51], 2010.

Case-mix variation can be taken into account by combining all relevant patient characteristics, that impact outcome and may differ between groups, in regression models ("case-mix adjustment models").[51] Case-mix adjustment, a cornerstone of health services research, is essential to make groups (e.g. treatment groups, hospitals) comparable on variables predicting the outcome of interest.[53] Case-mix adjustment has been applied to medical outcomes[54,55], PROMs[56,57], and clinical process measures[58,59]. Although the case-mix adjustment of traditional outcomes (e.g. mortality) has been well established, knowledge gaps still exist concerning the application of case-mix adjustment for PROMs among other measures.[5] Further empirical research is required to fill in these gaps, considering the ongoing debate about the appropriate use and advantages of case-mix adjustment.[5]

For fair inter-hospital comparisons, case-mix adjustment is not only dependent on age, sex, socio-economic status but on the clinical complexity of patients as well.[60] An example of clinical complexity is the presence of multi-morbidity or (co)morbidity. A variety of data sources to ascertain (co)morbidity have been described including patient-reported measures, health records, and administrative data.[61] However, there is still no consensus on the best data source for this important predictor.[62] Patient-reported (co)morbidity measures are easily administered with limited costs, and are a less labor-intensive data collection method than review of health records.[63,64] Despite these advantages, there is considerable skepticism about the accuracy of patient-reported morbidity measures.[65]

IMPLEMENTATION CHALLENGES

Besides methodological challenges in the analysis of PROMs, there are numerous practical, attitudinal and technical challenges in regards to the implementation and practice of routinely collecting PROMs in a clinical setting.[66] These challenges in practical application can be linked to various key components of the PROM implementation process: from selection of appropriate instruments, standardization in terms of frequency and method of PROM collection and review, IT-infrastructure, training and support for healthcare professionals, workload implications, registration burden, data security, and availability of resources. To date, studies on the implementation of PROMs have been scarce and limited to particular healthcare areas, e.g. palliative care.[67-69] As only a handful of health care institutions have been able to successfully implement routine PROM collection [70], it is necessary that more evidence is acquired on the implementation process. This includes the identification of "real world" facilitators and barriers, experienced

by those involved (e.g. patients, clinicians, nurse(s) (practitioners), administrative and IT-personnel) in the process. Such evidence may then guide the future endeavors of other healthcare organizations.[71]

EXAMPLES OF CLINICAL CARE ADDRESSED IN THIS THESIS: STROKE & BREAST CANCER

Quantitative methodology regarding the case-mix adjustment of PROMs, is the focus of research in the first part of this thesis. Studies are applied to data from acute ischemic stroke patients. Studies on the implementation and practical application of PROMs in clinical care, described in the second part of this thesis, focus mainly on breast cancer. Although ischemic stroke and breast cancer are two vastly different diseases in terms of etiology, setting (ambulatory care for stroke vs. "chronic" care for breast cancer), treatment, follow-up and prognosis, these two diseases rank among the top ten diseases responsible for premature death in the Netherlands.[72] Most studies in this thesis only include study populations in a hospital setting. Two reasons for this are the limited use of PROMs (and thus limited available data) in Dutch primary healthcare settings[73] and, more importantly, that the primary treatment of aforementioned diseases takes place in hospitals. As stroke and breast cancer represent different clinical settings, the aforementioned studies enable the reflection on the generalizability of both methodological and practical challenges of PROMs.

PROMS IN STROKE CARE

Stroke can be defined as an acute neurological deficit in localized area of the central nervous system due to a vascular cause. With over 5.5 million deaths per year, stroke is the second leading cause of mortality worldwide.[74] Stroke can be broadly categorized in acute ischemic stroke (approximately 85% of cases) or hemorrhagic stroke (including cerebral and subarachnoid hemorrhage).[75] Acute ischemic stroke is usually caused by a thrombosis or embolism that occludes a cerebral vessel, which results in a loss of neurological function.[76] In developed countries, stroke is also the leading cause of long-term disability (limitation of activities) which can encompass impairments in speech/ language (e.g. aphasia), physical functioning (e.g. mobility, gait, balance), cognitive functioning (e.g. memory, concentration), and psychosocial functioning (e.g. depression).[76-79] A decline in health-related quality of life (HRQOL) in stroke survivors has also been demonstrated, which

has been associated with age, stroke severity and functional status, among other predictors.[80]

Traditionally, clinical outcomes such as mortality, stroke recurrence and functional status (e.g. modified Ranking Scale (mRS), National Institute of Health Stroke Scale (NIHSS)) are considered important to quantify the impact of stroke on patients. In recent years, PROMs have become the gold standard in assessing patients' health status, and have been implemented widely in clinical stroke care with the goal of advocating shared decision-making, improving quality of stroke care, and benchmarking stroke centers.[81] A considerable number of PROMs, both generic (e.g. EQ-5D[82]) and disease-specific (e.g. Neuro-QOL (Quality of Life in Neurological Disorders)[83], SIS (Stroke Impact Scale)[84], SS-QOL (Stroke-Specific Quality of Life)[28]), are commonly used within clinical stroke care.[81]

PROMS IN BREAST CANCER CARE

Breast cancer is the most frequently diagnosed cancer amongst women worldwide. In the Netherlands, 1 in 7 women will be diagnosed with breast cancer.[85] As survival rates of (early-stage) breast cancer patients continue to increase due to advancements in screening, diagnostics and treatment, more attention is being paid to quality of life of this particular patient population.[86,87] Breast cancer can affect the (health-related) quality of life, from the time of diagnosis through the various treatment stages (e.g. chemotherapy) to the end of follow-up.

Although randomized trials have shown comparable survival rates between breast-conserving therapy (BCT) and mastectomy, there still remains a substantial knowledge gap about the relative impact of these various surgical treatment options on patients' quality of life and satisfaction with care.[88] Patient-reported outcome measures may provide a deeper understanding on important quality-of-life aspects following different (surgical) treatment strategies, which can guide the decision-making process in current and future patients.[88] Current research also focuses on the role of normative and reference PROM scores for enhancing shared decision-making between patients and healthcare providers.[38,89,90] Commonly used PROMs in both breast cancer research and clinical care include generic (e.g. SF-36[26]), disease-specific (e.g. EORTC-QLQ-Breast Cancer Module (EORTC-QLQ-BR23)[91]) and surgery-specific (e.g. BREAST-Q[92]) instruments. Despite a plethora of developed PROMs within the breast cancer population, there are only a few instruments that have sufficient psychometric evidence to support their use.[93]

AIMS AND OUTLINE OF THIS THESIS

This dissertation aims to study methodological aspects of analyzing patient-reported outcome measures (PROMs) and to provide insights into the implementation and practical application of PROMs in clinical care. The first part of this thesis focuses on case-mix (adjustment) for PROMs, while the second part addresses the implementation and practical application of PROMs.

Specifically, the questions that will be addressed in this thesis are:

1. Which general patient-reported morbidity instruments (in order to use for case-mix adjustment) have been published? (*Chapter 2*)
2. a) Which predictors are important for case-mix adjustment models for a generic patient-reported outcome measure (EQ-5D) within acute ischemic stroke care? (*Chapter 3*)
   b) How do case-mix adjustment models for the EQ-5D differ from models for clinical outcomes (mortality, modified Rankin Scale) within acute ischemic stroke care? (*Chapter 3*)
3. a) What are the differences in case-mix adjustment models between individual domains of a generic patient-reported outcome measure (EQ-5D) within acute ischemic stroke? (*Chapter 4*)
   b) What is the influence of case-mix adjustment on the reliability of individual EQ-5D domains as quality indicators for between-hospital comparisons in the quality of ischemic stroke care? (*Chapter 4*)
4. a) How have patient-reported outcome measures been administered within clinical breast cancer care? (*Chapter 5*)
   b) What is the impact of (routine) PROM collection on patients, care providers, and care processes? (*Chapter 5*)
   c) Which facilitators and barriers of PROM collection have been reported? (*Chapter 5*)
5. What is the course of PROM scores over time within (surgical) treatment strategies for breast cancer? (*Chapter 6*)
6. What are the normative BREAST-Q scores in a sample of Dutch women without a prior history of breast cancer? (*Chapter 7*)
7. What are the facilitators and barriers for implementing PROMs in clinical care, as reported by healthcare professionals of an academic center? (*Chapter 8*)

These aforementioned questions will be explored in two parts of this dissertation.

The first part of this thesis will study methods for case-mix adjustment of patient-reported outcome measures with its further implications for between-hospital comparisons. These methods will be applied to patient populations with acute ischemic stroke. It will also address patient-reported morbidity instruments that register this important case-mix factor for future adjustments.

In chapter 2, an overview of general patient-reported morbidity instruments including their psychometric properties is provided in a systematic review.

In chapter 3, the development and comparisons of three case-mix adjustment models for mortality, a functional outcome, and a generic patient-reported outcome measure is presented.

In chapter 4, between-hospital differences in patient-reported outcome measures after acute ischemic stroke are studied and it is assessed to what extent these differences are influenced by case-mix factors.

The second part of this thesis explores how patient-reported outcome measures can and have been implemented in clinical care, including the facilitators and barriers reported by healthcare professionals. These topics have been applied mostly to breast cancer care.

In chapter 5, a systematic review is presented on how patient-reported outcome measures have been routinely captured or administered in clinical breast cancer care, and what their impact has been on patients, clinicians and care processes.

In chapter 6, the course of prospectively collected patient-reported outcome measurement scores over time is demonstrated within several breast cancer surgical cohorts.

In chapter 7, the normative BREAST-Q data of Dutch women without a prior history of breast cancer is estimated and compared to internationally published data.

In chapter 8, a cross-sectional study is presented on the facilitators and barriers for implementing patient-reported outcome measurements in one academic medical center, as reported by surveyed healthcare professionals.

An overview of the included studies, the methodology and used data sources, is presented in Table 1.

Table 1 | Overview of Included Articles in This Thesis.

Chapter	Study title	Disease	Study type	Data source
2	Patient-Reported Morbidity Instruments: a Systematic Review	N/A	Systematic review	N/A
3	Value-based Healthcare in Ischemic Stroke Care: Case-Mix Adjustment Models for Clinical and Patient-Reported Outcomes	Ischemic Stroke	Empirical (Retrospective)	Sample population of the Dutch National Stroke Registry. Patients (N=1022) treated between 2014 and 2016.
4	Case-Mix Adjustment Models for Hospital Comparisons in Individual EQ-5D Domains after Acute Ischemic Stroke: a Substudy of the MR CLEAN Trial	Ischemic Stroke	Empirical (Prospective)	MR CLEAN (Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands), a multi-center randomized clinical trial comparing intra-arterial treatment (IAT) plus usual care (intervention group) with usual care alone (control group). Patients (N=500) were enrolled between 2010 and 2014.
5	Implementing Patient-Reported Outcome Measures in Clinical Breast Cancer Care: A Systematic Review	Breast Cancer	Systematic Review	N/A
6	A Prospective Analysis of Short-Term Patient-Reported Outcomes After Breast-Conserving Therapy and Mastectomy Without Reconstruction	Breast Cancer	Empirical (Prospective)	Patients (N=213) treated in the Erasmus University Medical Center, an academic hospital, between 2015 and 2019.
7	Normative BREAST-Q Data From A Dutch Sample of 9059 Women	N/A	Empirical (Cross-sectional)	Dutch women (N=9059) surveyed between February – July 2020.
8	Facilitators and Barriers of Implementing Patient Reported Outcome Measures in Routine Clinical Care: an Academic Center's Initial Experience	N/A	Empirical (Cross-sectional)	Care providers (N=61) at the Erasmus University Medical Center, an academic hospital, surveyed between January – March 2020.

With this series of aforementioned studies, an answer will be provided on the posed questions.

In chapter 9, a reflection on the studies in this thesis is provided with a discussion on how the use of PROMs can be improved through appropriate case-mix adjustment and practical application in clinical care. Summaries in English and Dutch conclude this thesis.

REFERENCES

1. Porter ME, Teisberg EO. *Redefining health care: creating value-based competition on results.* Harvard Business Press; 2006.
2. Porter ME, Lee TH. From Volume to Value in Health Care: The Work Begins. *Jama.* 2016;316(10):1047-1048.
3. Porter ME, and Thomas H. Lee. "The Strategy That Will Fix Health Care." Harvard Business Review 91, no. 10 (October 2013): 50–70.
4. Porter ME. What is value in health care? *N Engl J Med.* 2010;363(26):2477-2481.
5. Paddison C, Elliott M, Parker R, et al. Should measures of patient experience in primary care be adjusted for case mix? Evidence from the English General Practice Patient Survey. *BMJ quality & safety.* 2012;21(8):634-640.
6. *Nederlandse Federatie van UMC's (NFU). Bouwstenen voor werken aan waardegedreven zorg. . Date Accessed: 19-08-2020.*
7. *Zorginstituut Nederland (ZIN). Meer patiëntregie door meer uitkomstinformatie in 2022. https://www.zorginstituutnederland.nl/actueel/nieuws/2018/08/14/meer-patientregie-door-meer-uitkomstinformatie-in-2022 - :~:text=In%20het%20rapport%20'Meer%20pati%C3%ABntregie, ziektelast%20in%20Nederland%20te%20hebben-Date Accessed: 08-2020.*
8. Chun J, Bafford AC. History and background of quality measurement. *Clin Colon Rectal Surg.* 2014;27(1):5-9.
9. Kudzma EC. Florence Nightingale and healthcare reform. *Nurs Sci Q.* 2006;19(1):61-64.
10. Donabedian A. The end results of health care: Ernest Codman's contribution to quality assessment and beyond. *Milbank Q.* 1989;67(2):233-256; discussion 257-267.
11. Salem D. *Quality Measures The Revolution in Patient Safety and Outcomes: The Revolution in Patient Safety and Outcomes.* 2020.
12. Donabedian A. Evaluating the quality of medical care. 1966. *Milbank Q.* 2005;83(4):691-729.
13. Eliot JL, Peter F, Brian KR. Quality Measurement in Healthcare. *Annual Review of Medicine.* 2013;64(1):485-496.
14. http://www.lmcforum.org/wp-content/uploads/2016/09/Van-Eenennaam-Fred_Value_based-healthcare-in-europe_what-is-next.pdf.
15. Black N. Patient reported outcome measures could help transform healthcare. *Bmj.* 2013;346:f167.
16. Black N, Jenkinson C. Measuring patients' experiences and outcomes. *Bmj.* 2009;339:b2495.
17. Weldring T, Smith SMS. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health Serv Insights.* 2013;6:61-68.
18. Valderas JM, Alonso J, Guyatt GH. Measuring patient-reported outcomes: moving from clinical trials into clinical practice. *Med J Aust.* 2008;189(2):93-94.
19. McKenna SP. Measuring patient-reported outcomes: moving beyond misplaced common sense to hard science. *BMC Med.* 2011;9:86-86.
20. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. *BJA Education.* 2017;17(4):137-144.
21. Alrubaiy L, Hutchings HA, Williams JG. Assessing patient reported outcome measures: A practical guide for gastroenterologists. *United European Gastroenterol J.* 2014;2(6):463-470.
22. Devlin NJ, Appleby J, Buxton M, et al. *Getting the Most Out of PROMs: Putting Health Outcomes at the Heart of NHS Decision-Making.* King's Fund; 2010.

23. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of life research.* 2010;19(4):539-549.
24. Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use.* Oxford University Press, USA; 2015.
25. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med.* 2001;33(5):337-343.
26. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical care.* 1992;30(6):473-483.
27. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol.* 2010;63(11):1179-1194.
28. Williams LS, Weinberger M, Harris LE, Clark DO, Biller J. Development of a stroke-specific quality of life scale. *Stroke.* 1999;30(7):1362-1369.
29. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute.* 1993;85(5):365-376.
30. P MDF. *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes.* 2nd edition ed: Chichester: Wiley; 2013.
31. Langley GB, Sheppeard H. The visual analogue scale: its use in pain measurement. *Rheumatol Int.* 1985;5(4):145-148.
32. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica.* 1983;67(6):361-370.
33. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. *Arch Neurol.* 1989;46(10):1121-1123.
34. Riis CL, Bechmann T, Jensen PT, Coulter A, Steffensen KD. Are patient-reported outcomes useful in post-treatment follow-up care for women with early breast cancer? A scoping review. *Patient Relat Outcome Meas.* 2019;10:117-127.
35. Basch E, Deal AM, Kris MG, et al. Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment: A Randomized Controlled Trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2016;34(6):557-565.
36. Girgis A, Durcinoska I, Levesque JV, et al. eHealth System for Collecting and Utilizing Patient Reported Outcome Measures for Personalized Treatment and Care (PROMPT-Care) Among Cancer Patients: Mixed Methods Approach to Evaluate Feasibility and Acceptability. *Journal of medical Internet research.* 2017;19(10):e330.
37. Blackwell LS, Marciel KK, Quittner AL. Utilization of patient-reported outcomes as a step towards collaborative medicine. *Paediatric respiratory reviews.* 2013;14(3):146-151.
38. van Egdom LSE, de Kock MA, Apon I, et al. Patient-Reported Outcome Measures may optimize shared decision-making for cancer risk management in BRCA mutation carriers. *Breast cancer (Tokyo, Japan).* 2019.
39. Hilarius DL, Kloeg PH, Gundy CM, Aaronson NK. Use of health-related quality-of-life assessments in daily clinical oncology nursing practice: a community hospital-based intervention study. *Cancer.* 2008;113(3):628-637.

40. Damman OC, Jani A, de Jong BA, et al. The use of PROMs and shared decision-making in medical encounters with patients: An opportunity to deliver value-based health care to patients. *Journal of Evaluation in Clinical Practice.* 2020;26(2):524-540.
41. Van Der Wees PJ, Nijhuis-Van Der Sanden MW, Ayanian JZ, Black N, Westert GP, Schneider EC. Integrating the use of patient-reported outcomes for both clinical practice and performance measurement: views of experts from 3 countries. *The Milbank Quarterly.* 2014;92(4):754-775.
42. Wu AW, Kharrazi H, Boulware LE, Snyder CF. Measure once, cut twice—adding patient-reported outcome measures to the electronic health record for comparative effectiveness research. *Journal of clinical epidemiology.* 2013;66(8):S12-S20.
43. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. *BMC health services research.* 2013;13:211.
44. Howell D, Molloy S, Wilkinson K, et al. Patient-reported outcomes in routine cancer clinical practice: a scoping review of use, impact on health outcomes, and implementation factors. *Annals of oncology : official journal of the European Society for Medical Oncology.* 2015;26(9):1846-1858.
45. King S, Exley J, Parks S, et al. The use and impact of quality of life assessment tools in clinical care settings for cancer patients, with a particular emphasis on brain cancer: insights from a systematic review and stakeholder consultations. *Qual Life Res.* 2016;25(9):2245-2256.
46. Kotronoulas G, Kearney N, Maguire R, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2014;32(14):1480-1501.
47. Jensen RE, Rothrock NE, DeWitt EM, et al. The role of technical advances in the adoption and integration of patient-reported outcomes in clinical care. *Medical care.* 2015;53(2):153-159.
48. Rose M, Bezjak A. Logistics of collecting patient-reported outcomes (PROs) in clinical practice: an overview and practical examples. *Qual Life Res.* 2009;18(1):125-136.
49. Baumhauer JF. Patient-Reported Outcomes - Are They Living Up to Their Potential? *The New England journal of medicine.* 2017;377(1):6-9.
50. Smith PC, Mossialos E, Leatherman S, European Observatory on Health S, Policies, Papanicolas I. *Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects.* Cambridge University Press; 2009.
51. Lingsma H. *Measuring quality of care: Methods and applications to acute neurological diseases* [Ph.D. thesis], Erasmus University Rotterdam; 2010.
52. Zaslavsky AM. Statistical issues in reporting quality data: small samples and casemix variation. *International journal for quality in health care : journal of the International Society for Quality in Health Care.* 2001;13(6):481-488.
53. Mayo NE, Nadeau L, Levesque L, Miller S, Poissant L, Tamblyn R. Does the addition of functional status indicators to case-mix adjustment indices improve prediction of hospitalization, institutionalization, and death in the elderly? *Medical care.* 2005;43(12):1194-1202.
54. Lingsma HF, Steyerberg EW, Eijkemans MJ, Dippel DW, Scholte Op Reimer WJ, Van Houwelingen HC. Comparing and ranking hospitals based on outcome: results from The Neth-

erlands Stroke Survey. *QJM : monthly journal of the Association of Physicians.* 2010;103(2):99-108.

55. Beck N, Hoeijmakers F, van der Willik EM, et al. National Comparison of Hospital Performances in Lung Cancer Surgery: The Role of Case Mix Adjustment. *The Annals of thoracic surgery.* 2018;106(2):412-420.
56. Burgess R, Bishop A, Lewis M, Hill J. Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature. *Physiotherapy.* 2019;105(2):137-146.
57. Teale E, Young J, Dennis M, Sheldon T. Predicting patient-reported stroke outcomes: a validation of the six simple variable prognostic model. *Cerebrovascular diseases extra.* 2013;3(1):97-102.
58. Mehta RH, Liang L, Karve AM, et al. Association of patient case-mix adjustment, hospital process performance rankings, and eligibility for financial incentives. *Jama.* 2008;300(16):1897-1903.
59. Mesterton J, Lindgren P, Ekenberg Abreu A, et al. Case mix adjustment of health outcomes, resource use and process indicators in childbirth care: a register-based study. *BMC pregnancy and childbirth.* 2016;16(1):125.
60. Carey IM, Shah SM, Harris T, DeWilde S, Cook DG. A new simple primary care morbidity score predicted mortality and better explains between practice variations than the Charlson index. *Journal of Clinical Epidemiology.* 2013;66(4):436-444.
61. Klabunde CN, Harlan LC, Warren JL. Data Sources for Measuring Comorbidity: A Comparison of Hospital Records and Medicare Claims for Cancer Patients. *Medical care.* 2006;44(10):921-928.
62. Sarfati D. Review of methods used to measure comorbidity in cancer populations: No gold standard exists. *Journal of Clinical Epidemiology.* 2012;65(9):924-933.
63. Marrie RA, Yu BN, Leung S, et al. The Utility of Administrative Data for Surveillance of Comorbidity in Multiple Sclerosis: A Validation Study. *Neuroepidemiology.* 2013;40(2):85-92.
64. Skinner KM, Miller DR, Lincoln E, Lee A, Kazis LE. Concordance between respondent self-reports and medical records for chronic conditions: experience from the Veterans Health Study. *J Ambul Care Manage.* 2005;28(2):102-110.
65. Newell SA, Girgis A, Sanson-Fisher RW, Savolainen NJ. The accuracy of self-reported health behaviors and risk factors relating to cancer and cardiovascular disease in the general population: a critical review. *Am J Prev Med.* 1999;17(3):211-229.
66. Greenhalgh J, Meadows K. The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literature review. *Journal of Evaluation in Clinical Practice.* 1999;5(4):401-416.
67. Antunes B, Harding R, Higginson IJ, Euroimpact. Implementing patient-reported outcome measures in palliative care clinical practice: a systematic review of facilitators and barriers. *Palliative medicine.* 2014;28(2):158-175.
68. Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. *BMJ quality & safety.* 2014;23(6):508-518.
69. Bausewein C, Simon ST, Benalia H, et al. Implementing patient reported outcome measures (PROMs) in palliative care--users' cry for help. *Health and quality of life outcomes.* 2011;9:27-27.

70. Basch E. Patient-Reported Outcomes - Harnessing Patients' Voices to Improve Clinical Care. *The New England journal of medicine.* 2017;376(2):105-108.
71. Foster A, Croot L, Brazier J, Harris J, O'Cathain A. The facilitators and barriers to implementing patient reported outcome measures in organisations delivering health related services: a systematic review of reviews. *J Patient Rep Outcomes.* 2018;2:46-46.
72. http://www.healthdata.org/netherlands. http://www.healthdata.org/netherlands. Accessed August 18, 2020.
73. Weenink J-W, Braspenning J, Wensing M. Patient reported outcome measures (PROMs) in primary care: an observational pilot study of seven generic instruments. *BMC Family Practice.* 2014;15(1):88.
74. https://www.world-stroke.org/assets/downloads/Annual_Report_2018_online_fnal_COMPRESSED.pdf.
75. Musuka TD, Wilton SB, Traboulsi M, Hill MD. Diagnosis and management of acute ischemic stroke: speed is critical. *CMAJ.* 2015;187(12):887-893.
76. Phipps MS, Cronin CA. Management of acute ischemic stroke. *Bmj.* 2020;368:l6983.
77. Berthier ML. Poststroke Aphasia. *Drugs & Aging.* 2005;22(2):163-182.
78. van Rijsbergen MWA, Mark RE, de Kort PLM, Sitskoorn MM. Subjective Cognitive Complaints after Stroke: A Systematic Review. *Journal of Stroke and Cerebrovascular Diseases.* 2014;23(3):408-420.
79. Robinson RG, Jorge RE. Post-Stroke Depression: A Review. *Am J Psychiatry.* 2016;173(3):221-231.
80. Carod-Artal FJ. Determining quality of life in stroke survivors. *Expert Rev Pharmacoecon Outcomes Res.* 2012;12(2):199-211.
81. Reeves M, Lisabeth L, Williams L, et al. Patient-Reported Outcome Measures (PROMs) for Acute Stroke: Rationale, Methods and Future Directions. *Stroke.* 2018;49(6):1549-1556.
82. Golicki D, Niewada M, Buczek J, et al. Validity of EQ-5D-5L in stroke. *Qual Life Res.* 2015;24(4):845-850.
83. Cella D, Lai JS, Nowinski CJ, et al. Neuro-QOL: brief measures of health-related quality of life for clinical research in neurology. *Neurology.* 2012;78(23):1860-1867.
84. Duncan PW, Wallace D, Lai SM, Johnson D, Embretson S, Laster LJ. The stroke impact scale version 2.0: evaluation of reliability, validity, and sensitivity to change. *Stroke.* 1999;30(10):2131-2140.
85. van Der Waal D, Verbeek ALM, Den Heeten GJ, Ripping TM, Tjan-Heijnen VCG, Broeders MJM. Breast cancer diagnosis and death in the Netherlands: a changing burden. *The European Journal of Public Health.* 2015;25(2):320-324.
86. Bodai BI, Tuso P. Breast cancer survivorship: a comprehensive review of long-term medical issues and lifestyle recommendations. *Perm J.* 2015;19(2):48-79.
87. Fayanju OM, Mayo TL, Spinks TE, et al. Value-Based Breast Cancer Care: A Multidisciplinary Approach for Defining Patient-Centered Outcomes. *Ann Surg Oncol.* 2016;23(8):2385-2390.
88. Chen CM, Cano SJ, Klassen AF, et al. Measuring quality of life in oncologic breast surgery: a systematic review of patient-reported outcome measures. *Breast J.* 2010;16(6):587-597.
89. Lagendijk M, van Egdom LSE, Richel C, et al. Patient reported outcome measures in breast cancer patients. *Eur J Surg Oncol.* 2018;44(7):963-968.
90. Mundy LR, Homa K, Klassen AF, Pusic AL, Kerrigan CL. Breast Cancer and Reconstruction: Normative Data for Interpreting the BREAST-Q. *Plast Reconstr Surg.* 2017;139(5):1046e-1055e.

91. Sprangers MA, Groenvold M, Arraras JI, et al. The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: first results from a three-country field study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 1996;14(10):2756-2768.
92. Pusic AL, Klassen AF, Scott AM, Klok JA, Cordeiro PG, Cano SJ. Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. *Plast Reconstr Surg.* 2009;124(2):345-353.
93. Kanatas A, Velikova G, Roe B, et al. Patient-reported outcomes in breast oncology: a review of validated outcome instruments. *Tumori.* 2012;98(6):678-688.

PART I

Methodological Aspects of PROMs

Chapter 2 Patient-Reported Morbidity Instruments: A Systematic Review

Arvind Oemrawsingh, MD MHS[1], Nishwant Swami, BA[2]; José M. Valderas, MD PhD MPH[3]; Jan A. Hazelzet, MD PhD[1]; Andrea L. Pusic MD MHS FACS FRCSC[4]; Richard E. Gliklich, MD[5]; Regan W. Bergmark, MD[5,6]

[1] *Center for Medical Decision Making, Department of Public Health, Erasmus University Medical Center Rotterdam, PO Box 2040, 3000 CA, Rotterdam, the Netherlands*
[2] *University of Massachusetts Medical School, Worcester, MA, USA*
[3] *International Society for Quality of Life Research (ISOQOL), Health Services & Policy Research, University of Exeter Medical School, Exeter, England, United Kingdom*
[4] *Division of Plastic and Reconstructive Surgery, Patient Reported Outcomes, Value and Experience (PROVE) Center, Brigham and Women's Hospital, Boston, MA, USA*
[5] *Department of Otolaryngology – Head and Neck Surgery, Harvard Medical School, Boston, MA, USA*
[6] *Patient Reported Outcomes, Value and Experience (PROVE) Center, Brigham and Women's Hospital, Boston, MA, USA*

Value in Health, 2020 Jun;23(6):791-811.

ABSTRACT

Objectives

Although comorbidities play an essential role in risk adjustment and outcomes measurement, there is little consensus regarding the best source of this data. The aim of this study was to identify general patient-reported morbidity instruments and their measurement properties.

Methods

A systematic review was conducted using multiple electronic databases (Embase, Medline, Cochrane Central, and Web of Science) from inception to March 2018. Articles focusing primarily on the development or subsequent validation of a patient-reported morbidity instrument were included. After including relevant articles, the measurement properties of each morbidity instrument were extracted by 2 investigators for narrative synthesis.

Results

A total of 1005 articles were screened, of which 34 eligible articles were ultimately included. The most widely assessed instruments were the Self-Reported Charlson Comorbidity Index (n = 7), the Self-Administered Comorbidity Questionnaire (n = 3), and the Disease Burden Morbidity Assessment (n = 3). The most commonly included conditions were diabetes, hypertension, and myocardial infarction. Studies demonstrated substantial variability in item-level reliability versus the gold standard medical record review (κ range 0.66-0.86), meaning that the accuracy of the self-reported comorbidity data is dependent on the selected morbidity.

Conclusions

The Self-Reported Charlson Comorbidity Index and the Self-Administered Comorbidity Questionnaire were the most frequently cited instruments. Significant variability was observed in reliability per comorbid condition of patient-reported morbidity questionnaires. Further research is needed to determine whether patient-reported morbidity data should be used to bolster medical records data or serve as a stand-alone entity when risk adjusting observational outcomes data.

Keywords

Comorbidity; Health Services; Morbidity; Patient Report; Psychometrics; Self-Report; Surveys and Questionnaires

INTRODUCTION

Value-based healthcare (VBHC) initiatives rely on risk adjustment to compare patient populations across hospitals. In addition to understanding the index disease of interest, comorbid conditions are necessary for case-mix adjustment. "Morbidity" is defined here as the presence of medical conditions. Clinicians are increasingly grappling with the challenges of treating patients with multiple co-occurring diseases (multi-morbidity). In addition to treatment difficulties, multi-morbidity is often associated with worse outcomes including decreased quality of life, psychological distress, longer hospital stays, more postoperative complications, higher cost of care, and higher mortality.[1, 2]

To identify opportunities for outcomes improvement, registries and groups like the International Consortium for Healthcare Outcomes Measurement (ICHOM) have attempted to standardize and compare observational data across hospitals.[3] These comparative studies often rely on risk adjustment algorithms to account for clinical differences in patient populations.[4] In analyzing a changing medical landscape with more multi-morbidity patients, reliable morbidity data has increasingly become a focal point for fair benchmarking as part of the shift toward VBHC.

Accurate inclusion of comorbidities in large data sets has proven to be a vexing problem. Although morbidity plays a crucial role in risk adjustment, risk stratification, and outcomes measurement, there is little consensus regarding the best source of this data. Comparisons of morbidity data from different sources have displayed significant variations.[5] Notable inconsistencies have been observed when morbidity data is collected from administrative sources, such as claims data.[6] Administrative data generally underreports comorbid conditions, leading to a lack of accounting for overall level of sickness of the patient.[7-9] Although some studies have shown more accurate information in hospital chart reviews, concerns arise regarding the burden of collection and the feasibility of wide-scale use.[10, 11]

To obtain more accurate morbidity data feasibly, clinicians have increasingly turned to patient-reported instruments as a potential alternative.[12, 13] The objective of this study was to provide a comprehensive evidence base of validated patient-reported morbidity instruments to aid in the selection of these instruments for use in clinical practice. Although many disease-specific morbidity instruments exist, our study examined questionnaires applicable to the broader patient population to allow for broader implementation across a healthcare system.

METHODS

Design and Rationale

Risk adjustment for the comparison of outcomes data across international healthcare centers relies on the accurate capture of predictor variables such as extent of morbidity. Because standard outcome sets could be used among health institutions with no or different electronic medical record and administrative data structures, a systematic review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[14] of studies about the development or subsequent validation of self-reported comorbidity assessments.

Literature Search

An exhaustive search strategy was developed in Embase.com by a medical librarian experienced in systematic review searches.15 To retrieve articles about the validation of questionnaires on comorbidity, the search strategy combined thesaurus terms (Emtree terms for Embase and MeSH terms for Medline) with terms in the title or abstract for 3 elements: comorbidity, questionnaires, and validation or reliability.

The search strategy for Embase was optimized to find all potentially relevant terms and then translated to Medline (Ovid), Cochrane CENTRAL, and Web of Science Core Collection.[16] Additional references were retrieved from Google Scholar (the first 100 references as sorted by relevance), literature lists of relevant reviews, and included references. Abstracts needed to be in English, but there were no restrictions on the language of the manuscript or country of publication in the search strategy. The databases were last searched on March 5, 2018. The full search strategies for all the databases are included in the online in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.02.006.

Study Selection

The inclusion criterion for studies was a primary focus on the development or subsequent validation (e.g reliability and prediction/ association with outcomes) of an instrument for collecting information on the presence of morbidity directly from the patient.

The exclusion criteria were:

- lack of any methodological description of instrument development (validation and/ or reliability of instrument)
- description of the use of a patient-reported comorbidity questionnaire for risk-adjustment purposes or for deriving health utilities
- focus in the patient-reported morbidity instrument on a subset of specific conditions (based on nosologic criteria), thereby making the instrument not generalizable to a larger patient population (e.g. a list of mental health comorbidities for psychiatric patients)

The search results were deduplicated[17] and then imported into Covidence (www.covidence.org, Melbourne, Australia), a Cochrane technology web-based platform developed specifically to screen and track articles through the inclusion and exclusion criteria process of a systematic review. In Covidence, the titles and abstracts of each reference were independently screened for relevance by 2 reviewers. The screening phase was conducted in the following order:

1. The first screening was based on the title and abstract. In the event that the article's aim did not meet the inclusion criteria but nevertheless mentioned a patient-reported medical comorbidity questionnaire, the full text was reviewed to determine the instrument used and references were further screened for any potential missing articles on the comorbidity questionnaire.
2. The second screening was based on the full-text assessment of retained articles. Studies that still did not meet our inclusion criterion were subsequently removed.

If the authors had any disagreements on article eligibility during the first screening, the study in question would be screened in a full-text version. Consensus on the inclusivity of selected articles was ultimately reached by the authors.

Included Comorbidities

The number of comorbidities, as well as a list of included comorbid conditions, was evaluated for every survey instrument. Additional survey questions, such as those evaluating the condition severity (impact on daily activities/ functional status) or medication use for a comorbidity, were also noted.

Reliability

Measures of reliability (e.g. test-retest reliability; patient-report vs other data sources such as medical records, administrative data, or laboratory testing) at either item-level or overall instrument level were catalogued for all studies. Be-

cause of the anticipated heterogeneity in the reporting between the studies, both reliability-specific values (intra-class correlation coefficient [ICC] and Kappa [κ] values) and other measures of the morbidity instrument's performance (Spearman correlation coefficient, sensitivity/ specificity, and positive and negative predictive values) were included. The kappa values were measured based on the presence of the condition in the self-reported instrument versus the medical record (or administrative record), which was considered the gold standard. The articles mention that reporting a condition by self-report that is absent in the medical record could also imply a deficiency with the medical record. Kappa values > 0.80 indicate excellent agreement, 0.61 to 0.80 good agreement, 0.41 to 0.60 moderate agreement, 0.21 to 0.40 fair agreement, and < 0.20 poor agreement.[18] Spearman correlation coefficients are categorized as ≤ 0.20 (poor), 0.30 to 0.59 (fair), 0.60 to 0.79 (moderate), and > 0.80 (strong).[19]

Evaluated Outcomes

All instances where patient-reported morbidity instruments were used to assess association with or predict certain outcomes as part of the validation study were documented for this review. Examples of the outcome metrics included are mortality, disease response, patient-reported outcome measures (PROMs), adverse events and other events of interest, and healthcare utilization/ costs, as per the categories of outcomes used in the Agency for Healthcare Research and Quality (AHRQ) Outcome Measures Framework (OMF).[20]

Questionnaire Length, Duration, Responsiveness, and Utilization

The length (number of items/ questions) of the instrument and duration of completion was documented, when available, as was the route of administration (self-administration vs administration by a clinical or research associate). Finally, the number of times the paper had been cited in Web of Science was also noted.

RESULTS

Included Studies

Figure 1 details the search and inclusion strategy. In total, 1005 studies met our search criteria; 70 studies met our criteria for inclusion in the full-text assessment. Thirty-six studies were eliminated after the full-text review, leaving 34 studies for inclusion in this systematic review.

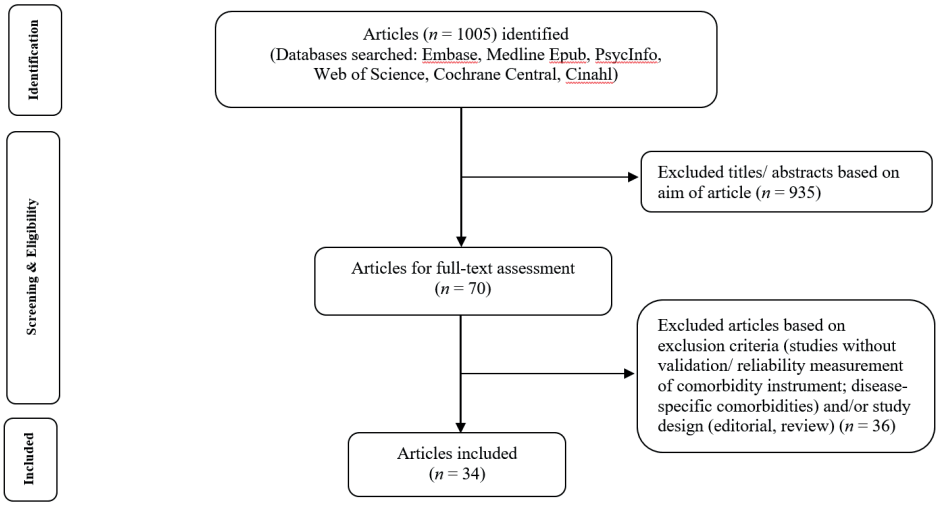


Figure 1 | Flowchart of Relevant Article Selection.

A summarized overview of all the included articles in this systematic review is included in Table 1. Descriptive characteristics, reliability, validity, and evaluated outcomes of morbidity instruments are shown in Table 2.[21-67]

Included Patient-Reported Morbidity Instruments

Ten original patient-reported morbidity instruments were identified, with most of these development studies being conducted in the United States. The instruments considered originals were: the Self-Reported Charlson Comorbidity Index (SR-CCI),[11] the Self-Administered Comorbidity Questionnaire (SCQ),[27] the Disease Burden Morbidity Assessment (DBMA),[32] the Comorbidity Symptom Scale (CmSS),[38] the Patient Self-Administered Health History Questionnaire,[47] the Multi-Morbidity Assessment Questionnaire for Primary Care (MAQ-PC),[59] the Patient-Based Comorbidity Index (CI),[64] the Health Impact Index (HII),[50] the Seattle Index of Comorbidity (SIC),[48] and an unnamed prognostic index (including comorbidities).[60] The SR-CCI and SCQ instruments were the most frequently cited. Other included articles were translation and cross-cultural adaptation studies, variations of these questionnaires (e.g. with a small number of items added or removed), or validation studies.

Table 1 | Summarized Overview of Included Studies.

Patient-Reported Morbidity Instruments	Selected article and country of origin	Availability of reliability data	Evaluated outcomes	Method of questionnaire administration	Time to questionnaire completion	Number of Citations
Patient-Reported Charlson Comorbidity Index	Katz, 1996*† United States	Item-level: + Overall: +	Mortality: - PROM: - Healthcare utilization: +	Self- or interviewer-administered	10 minutes	758
	Susser, 2008* Canada	Item-level: + Overall: +	Mortality: - PROM: + Healthcare utilization: -	Self-administered or filled out by proxy	-	19
	Corser, 2008* United States	Item-level: + Overall: +	Mortality: - PROM: - Healthcare utilization: -	Interviewer-administered	-	53
	Olomu, 2012* United States	Item-level: - Overall: -	Mortality: - PROM: + Healthcare utilization: -	Interviewer-administered	-	23
	Ng, 2015* Singapore	Item-level: + Overall: +	Mortality: - PROM: + Healthcare utilization: -	Self- or interviewer-administered	15 minutes	5
	Habbous, 2013* Canada	Item-level: + Overall: +	Mortality: + PROM: - Healthcare utilization: -	Self-administered	-	11
	Chaudhry, 2005* United States	Item-level: + Overall: -	Mortality: + PROM: - Healthcare utilization: +	Self-administered	1 minute	172
Self-Reported Comorbidity Questionnaire	Sangha, 2003*† United States	Item-level: + Overall: +	Mortality: - PROM: + Healthcare utilization: +	Self-administered	-	757
	Stolwijk, 2014* Netherlands/ Belgium	Item-level: + Overall: +	Mortality: - PROM: + Healthcare utilization: -	Self-administered	-	17
	Robinski, 2016* Germany	Item-level: + Overall: -	Mortality: - PROM: + Healthcare utilization: -	Self-administered	-	2

Table 1 | Summarized Overview of Included Studies. (continued)

Patient-Reported Morbidity Instruments	Selected article and country of origin	Availability of reliability data	Evaluated outcomes	Method of questionnaire administration	Time to questionnaire completion	Number of Citations
Disease Burden Morbidity Assessment	Bayliss, 2005† United States	Item-level: + Overall: -	Mortality: - PROM: + Healthcare utilization: -	Self-administered	-	133
	Poitras, 2012 Canada	Item-level: + Overall: -	Mortality: - PROM: - Healthcare utilization: -	Self-administered	<15 minutes	19
	Wijers, 2017 Spain	Item-level: - Overall: -	Mortality: - PROM: + Healthcare utilization: -	Self-administered	-	2
	Simpson, 2004 United States	Item-level: + Overall: -	Mortality: - PROM: - Healthcare utilization: -	Self-administered	-	214
Comorbidity Symptom Scale	Crabtree, 2000* England	Item-level: - Overall: -	Mortality: - PROM: + Healthcare utilization: -	Interviewer-administered	<10 minutes	29
	De-loyde, 2015 Australia	Item-level: - Overall: -	Mortality: - PROM: - Healthcare utilization: -	Self-administered	-	10
	Gad, 2012 United States	Item-level: + Overall: -	Mortality: - PROM: - Healthcare utilization: -	Self-administered	-	3
	Hansen, 2014 Germany	Item-level: + Overall: -	Mortality: - PROM: - Healthcare utilization: -	Interviewer-administered	-	-
	Horton, 2010 Canada/ United States	Item-level: + Overall: -	Mortality: - PROM: - Healthcare utilization: -	Self-administered	11 minutes (mean)	63
Questionnaire from CALAS study	Iecovich, 2013 Israel	Item-level: + Overall: +	Mortality: - PROM: - Healthcare utilization: -	Interviewer-administered	30 – 45 minutes	1

Table 1 | Summarized Overview of Included Studies. (continued)

Patient-Reported Morbidity Instruments	Selected article and country of origin	Availability of reliability data	Evaluated outcomes	Method of questionnaire administration	Time to questionnaire completion	Number of Citations
	Klabunde, 2006 United States	Item-level: - Overall: -	Mortality: - PROM: - Healthcare utilization: -	Self-administered	-	114
	Boissonnault, 2005 United States	Item-level: + Overall: +	Mortality: - PROM: - Healthcare utilization: -	Self-administered	-	15
Seattle Index of Co-morbidity	Fan, 2002* United States	Item-level: - Overall: -	Mortality: + PROM: + Healthcare utilization: +	Self-administered	-	123
Health Impact Index	Lorem, 2016* Norway	Item-level: - Overall: +	Mortality: - PROM: + Healthcare utilization: -	Self-administered	-	4
	Lucke, 2016 Germany	Item-level: + Overall: +	Mortality: - PROM: - Healthcare utilization: -	Self-administered	-	12
Cornell Medical Index	Md Yusof, 2010* United Kingdom	Item-level: - Overall: +	Mortality: + PROM: - Healthcare utilization: -	Self-administered	-	2
Questionnaire from CHOICE study	Merkin, 2007 United States	Item-level: + Overall: -	Mortality: - PROM: - Healthcare utilization: -	-	-	70
Questionnaire from AHEAD study	Mukerji, 2007 United States	Item-level: + Overall: +	Mortality: - PROM: - Healthcare utilization: -	Self-administered or interviewer administered	-	41
	Paleri, 2002* United Kingdom	Item-level: + Overall: -	Mortality: - PROM: - Healthcare utilization: -	Self-administered	8.3 minutes	22

Table 1 | Summarized Overview of Included Studies. (continued)

Patient-Reported Morbidity Instruments	Selected article and country of origin	Availability of reliability data	Evaluated outcomes	Method of questionnaire administration	Time to questionnaire completion	Number of Citations
Multi-Morbidity Assessment Questionnaire for Primary Care	Pati, 2016 India	Item-level: + Overall: -	Mortality: PROM: Healthcare utilization:	-	20 – 25 minutes	5
	Lee, 2006* United States	Item-level: - Overall: -	Mortality: + PROM: - Healthcare utilization: -	Interviewer-administered	-	416
Self-Report Comorbidity	Voaklander, 2004* Canada	Item-level: - Overall: +	Mortality: - PROM: + Healthcare utilization: +	-	-	12
(Patient-Based) Comorbidity Index	Selim, 2004* United States	Item-level: - Overall: -	Mortality: + PROM: + Healthcare utilization: +	Interviewer-administered	-	150
Questionnaire from LACE study	Vigen, 2016 United States	Item-level: + Overall: -	Mortality: + PROM: - Healthcare utilization: -	Self- or interviewer-administered	-	7

Note. + = described in study; – = unknown/ not described in study; * = instrument associated with an index; † = main developmental study
AHEAD, indicates Action for Health in Diabetes; CALAS, Cross-Sectional and Longitudinal Aging Study; CHOICE, Choices for Healthy Outcomes in Caring for End-stage renal disease study; LACE, Life After Cancer Epidemiology study; PROM, Patient-Reported Outcome Measures.

Table 2 | Detailed Overview of Included Studies.

	Items	Study Population (N)	Item-Level Reliability vs. Other Data Sources
Katz[11]	AIDS Any tumor Cerebrovascular disease Chronic pulmonary disease Congestive heart failure Connective tissue disease Dementia Diabetes (end-organ damage) Diabetes (mild to moderate) Hemiplegia Leukemia Liver disease Lymphoma Metastatic tumor Moderate/ Severe renal disease Myocardial infarction Peripheral vascular disease Ulcer disease	170 inpatients from 6 care units (3 medical and 3 surgical) at 1 hospital. **Characteristics:** Female 55% Mean age 65.3 years (± SD 8.8) Caucasian 82% College level or higher 50% Surgical 54%	**Comparison: Medical record-derived CCI** **Results:** Kappa: 0.35 (ulcer disease/ diabetes with end-organ damage) – 0.85 (leukemia) Sensitivity: - Specificity: - PPV: - NPV: - Agreement between self-reported CCI and medical record-derived CCI ranged from 83% (any tumor) to 100% (AIDS).
Susser[9]	As per Katz	520 elderly patients ready to be discharged home from the ER. Data from a previously published RCT[21]. **Characteristics:** Female 60% Age group >75 years 57%	**Comparison: Administrative data-derived CCI** **Results:** Kappa: Highest $\kappa = 0.55$ (chronic pulmonary disease). Individual Kappa values with range were not described for all conditions. Sensitivity: - Specificity: - PPV: - NPV: - Four conditions were reported more frequently by self-report (myocardial infarction, ulcer disease, diabetes with end-organ damage, connective tissue disease), while five (hemiplegia, mild-moderate diabetes, solid tumor, lymphoma, dementia) were more frequently reported in administrative data.
Corser[22]	As per Katz	525 patients admitted for acute coronary syndrome in 5 hospitals. **Characteristics:** Female 36.4% Mean age 59.73 years (± SD 12) Caucasian 84.4% College level or higher 43.8%	**Comparison: Medical record-derived CCI** **Results:** Kappa: 0.07 (CTD, RA) – 0.80 (diabetes). Only conditions with a prevalence of at least 3% (in each data source) were included in the Kappa analysis. Sensitivity: - Specificity: - PPV: - NPV: -

Overall Instrument Reliability vs. Other Data Sources	Evaluated Outcome(s)
Comparison: Medical record-derived CCI **Results:** Self-reported CCI score was higher versus medical-record derived CCI (1.99 ± SD 2.13 vs. 1.59 ± SD 1.80, $p < 0.01$) Spearman's r range 0.63 ($p = 0.0001$) (full index) – 0.70 ($p = 0.0001$) (when the solid tumor item was excluded from the analysis)	Measured in 49 and 56 patients on medical and surgical service respectively. **Results:** Mortality: - PROM: - Health care utilization: *Hospitalizations in last year:* Spearman's r range 0.17 – 0.31, $p < 0.05$ *Number of prescription medication*: Spearman's r range 0.26 – 0.44, $p < 0.05$ *Hospital charges during admission:* Spearman's r range 0.09 – 0.26, $p < 0.05$) *Length of stay:* Spearman's r range 0.15 – 0.20
Comparison: Administrative data-derived CCI **Results:** Poor agreement between self-reported and administrative data-derived CCI, indicated by an (overall) ICC = 0.43 (95% CI 0.40 – 0.47).	**Comparison: Administrative data-derived CCI** **Results:** Mortality: - PROM: *ADL (functional decline):* predictive ability of self-reported vs. administrative data-derived CCI was measured with unweighted (for sampling) AUC = 0.51 vs. AUC = 0.54 and weighted AUC = 0.54 vs. AUC = 0.50, $p>0.05$) Health care utilization: *Hospital days:* self-reported vs. administrative data-derived CCI was measured with unweighted AUC = 0.63 vs. AUC = 0.63 and weighted AUC = 0.68 vs. AUC = 0.69, $p > 0.05$) *ER visits:* unweighted AUC = 0.64 vs. AUC = 0.65 and weighted AUC = 0.67 vs. AUC = 0.63, $p > 0.05$)
Comparison: Medical record-derived CCI **Results:** Self-reported CCI (composite) scores were higher than medical record-derived CCI scores (mean 1.78 ± SD 1.99 vs. mean 1.27 ± SD 1.43). Correlation between self-reported and medical record-derived CCI composite scores were fair (Spearman's r = 0.57, $p < 0.01$).	-

Table 2 | Detailed Overview of Included Studies. (continued)

	Items	Study Population (N)	Item-Level Reliability vs. Other Data Sources
Olomu[12]	As per Katz	525 patients admitted for acute coronary syndrome in 5 hospitals. Data from a previously published RCT.[23] **Characteristics:** Female 36.4% Caucasian 84.4% College level or higher 43.8%	-
Ng[24]	As per Katz	301 rheumatic patients from 1 tertiary hospital. **Characteristics:** Female 61.5% Median age 51 years (21-79) Chinese 68.8% College level or higher 54.7%	**Comparison: Medical record-derived CCI** **Results:** Kappa: 0.189 (diabetes with end-organ damage) – 0.764 (diabetes). Kappa values was only calculated for 8/18 conditions that did not have any cell values of zero. Sensitivity: 33.3 (diabetes with end-organ damage) – 100% (myocardial infarction) Specificity: 58.9 (CTD, RA) – 99.1% (CVA) PPV: - NPV: - Agreement between self-reported CCI and medical record-derived CCI ranged from 74.1% (CTD/ RA) to 100% (leukemia, lymphoma, metastatic solid tumor, AIDS).

Overall Instrument Reliability vs. Other Data Sources	Evaluated Outcome(s)
-	**Comparison: Medical record-derived CCI.** **Results:** Mortality: - PROM: *ASI (functional capacity)*: Prediction at 3 months was slightly better with SCQ vs. the CCI (R^2 = 0.340, p < 0.0005 vs. R^2 = 0.331, p < 0.0035), while it was slightly better with CCI vs. SCQ (R^2 = 0.370, p < 0.0005 vs. R^2 = 0.358, p < 0.0005) at 8 months. *EQ-5D (health-related quality of life)*: Only the SCQ significantly predicted EQ-5D scores at 3 and 8 months (R^2 = 0.288 and R^2 = 0.265, p < 0.0005), whereas the CCI did not (R^2 = 0.262, p > 0.201 and R^2 = 0.245, p > 0.132). Health care utilization: -
Comparison: Medical record-derived CCI **Results:** Median self-reported composite CCI scores were higher than the medical record-derived CCI scores, indicating that conditions were generally reported more frequently by self-report than EHR review. Self-reported composite CCI scores had moderate agreement (ICC = 0.513, p < 0.001) and strong correlation (Spearman's r = 0.570, p < 0.001) with the medical record-derived CCI scores.	**Comparison: Medical record-derived CCI** **Results:** Mortality: - PROM: *SF-36 (health-related quality of life):* Self-reported CCI was negatively associated with PCS (β = -2.56, p < 0.001) and MCS (β = -1.24, p = 0.044). Medical record-derived CCI scores had a similar trend but coefficients didn't reach statistical significance. Health care utilization: -

Table 2 | Detailed Overview of Included Studies. (continued)

	Items	Study Population (N)	Item-Level Reliability vs. Other Data Sources
Habbous[25]	*Exposures* Smoking and alcohol *Conditions/Diseases* Chronic cough/bronchitis Dementia (e.g. Alzheimer's) Diabetes (eye/ kidney problems) (Past) Dialysis requirement Emphysema Heart failure Hemiplegia Hepatitis HIV/AIDS Liver disease Myocardial infarction Other joint/bone problems Past cancer history Peripheral vascular disease Rheumatoid arthritis Serious kidney problems Stomach ulcers (test-proven) Stroke/ Mini-stroke	882 head-and-neck cancer patients. **Characteristics:** Female 23% Median age 61.5 years (61-62.5) Caucasian 84%	**Comparison: Medical record-derived CCI** **Results:** Kappa: 0.16 (hemiplegia) – 0.93 (diabetes) Sensitivity: - Specificity: - PPV: - NPV: - Positive agreement between self-reported CCI and medical record-derived CCI ranged from 17 (hemiplegia) to 94% (diabetes). Negative agreement between self-reported CCI and medical record-derived CCI ranged from 84 (CTD) to 100% (dementia).

Overall Instrument Reliability vs. Other Data Sources	Evaluated Outcome(s)
Comparison: Medical record-derived CCI **Results:** Patient-reported CCI scores were higher than the medical record-derived CCI scores (mean 1.01 (95% CI 0.9 – 1.1) vs. 0.74 (95% CI 0.7 – 0.8), p < 0.0001). Comorbidities were reported more often by patients in comparison to medical records review. Overall agreement between patient-reported CCI and medical record-derived CCI was measured as κ=0.37, which improved if CTD (κ=0.52) or COPD (κ=0.43) was removed from the patient-reported CCI score.	**Comparison: Medical record-derived CCI** **Results:** Mortality: *Overall survival:* Both patient-reported CCI (HR 1.62 (95% CI 1.18 – 2.24), p = 0.003) and medical record-derived CCI (HR 1.97 (95% CI 1.38 – 2.80), p = 0.0002) were significantly associated with overall survival after multivariate (age, sex, marital status, stage and disease site) adjustment, when at least 2 comorbidities were present. PROM: - Health care utilization: -

Table 2 | Detailed Overview of Included Studies. (continued)

	Items	Study Population (N)	Item-Level Reliability vs. Other Data Sources
Chaudhry[26]	Asthma/ Emphysema/ Chronic bronchitis Arthritis or rheumatism Cancer (diagnosed within past 3 years) Diabetes Digestive problems (e.g. ulcer/ colitis/ gallbladder disease) Heart trouble (e.g. angina/ CHF/ CAD) HIV or AIDS Kidney disease Liver problems (cirrhosis) Stroke	7761 hospitalized general medicine patients at a single center. **Characteristics:** Female 62% Mean age 56-57 years African-American >80% MMSE score > 17	**Comparison: Administrative data-derived CCI** **Results:** Item-level data vs. one-year look-back: Kappa: 0.04 (stomach ulcer) – 0.83 (diabetes) Sensitivity: 44 (cancer) – 86% (diabetes, HIV/ AIDS) Specificity: 48 (arthritis or rheumatism) – 98% (HIV/ AIDS) PPV: 3 (stomach ulcers) – 90% (diabetes) NPV: 91 (asthma, emphysema or bronchitis) – 100% (HIV/ AIDS) Item-level data vs. index hospitalization: Kappa: 0.06 (arthritis or rheumatism) – 0.82 (diabetes) Sensitivity: 43 (cancer) – 91% (diabetes) Specificity: 47 (arthritis or rheumatism) – 95% (liver disease, cancer diagnosed in past 3 years) PPV: 9 (arthritis or rheumatism) – 84% (diabetes) NPV: 93 (asthma, emphysema or bronchitis) – 99% (heart disease, kidney disease, liver disease, cancer diagnosed in past 3 years) No statistically significant differences in Kappa values, sensitivities, specificities, or positive or negative predictive values was observed for 1-year look-back periods or index hospitalization.

Overall Instrument Reliability vs. Other Data Sources	Evaluated Outcome(s)
-	The predictive power of the *self-reported* CCI was constructed with 4 different logistic regression models performed in a validation cohort (N=3870). Model 1: age, sex + original CCI weight Model 2: age, sex + study-specific CCI weight Model 3: age, sex, diagnosis-related group weight + original CCI weight, Model 4: age, sex, diagnosis-related group weight + study-specific CCI weight **Results:** Mortality: *One-year mortality*: AUC's for the self-reported CCI were 0.70 (0.68 – 0.73) for model 1, 0.72 (0.70 – 0.75) for model 2, 0.75 (0.72 – 0.77) for model 3, and 0.76 (0.73 – 0.78) for model 4. AUC's were slightly less compared to the administrative data-derived CCI indices ($p < 0.001$). PROM: - Health care utilization: *Log total costs:* R^2 values for the different regression models ranged from 0.02 (models 1 & 2) – 0.33 (models 3 & 4) *Log length of stay:* R^2 values for the different regression models ranged from 0.01 (model 1) – 0.22 (models 3 & 4)

Table 2 | Detailed Overview of Included Studies. (continued)

	Items	Study Population (N)	Item-Level Reliability vs. Other Data Sources
Sangha[27]	Anemia or other blood disease Back pain Cancer Depression Diabetes Heart disease High blood pressure Kidney disease Liver disease Lung disease Osteoarthritis/ Degenerative arthritis Other medical problems (optional) Rheumatoid arthritis Ulcer or stomach disease	170 hospitalized patients from 6 care units at one hospital. **Characteristics:** Female 55% Mean age 65.3 years (± SD 8.8) Caucasian 82% College level or higher 50%	**Comparison: Medical record-derived CCI** **Results:** Kappa: 0.27 (lung disease) – 0.93 (liver disease) Sensitivity: - Specificity: - PPV: - NPV: - Overall agreement between the SCQ and the medical record-derived CCI ranged from 78% (heart disease) to 99% (liver disease).
Stolwijk[28]	Dutch modified version (mSCQ) of the SCQ instrument[27] Anemia or other blood disease Back pain Cancer Depression Diabetes Heart disease High blood pressure Kidney disease Liver disease Lung disease Osteoarthritis Other non-specified medical problems (optional; max. 3) Rheumatoid arthritis Ulcer or stomach disease	98 outpatients with ankylosing spondylitis. Data from the OASIS study.[29] **Characteristics:** Female 29.6% Mean age 53.9 years (± SD 11.4) College level or higher 15.7%	**Comparison: Medical records** **Results:** Kappa: 0.14 (osteoarthritis, ulcer disease) – 1.00 (cancer). Kappa analysis included 10 conditions. Sensitivity: - Specificity: - PPV: - NPV: -

Overall Instrument Reliability vs. Other Data Sources	Evaluated Outcome(s)
Comparison: Medical record-derived CCI **Results:** SCQ scores were higher than the medical record-derived CCI scores (mean 5.61 ± SD 4.1 vs. 1.59 ± SD 2.13) SCQ had a fair correlation (Spearman's r = 0.32) with the medical record-derived CCI, which slightly increased (Spearman's r = 0.55) when questionnaires were truncated to only comparable items.	**Comparison: Medical record-derived CCI** Results for medical patients: Mortality: - PROM: *SF-36 (health-related quality of life):* SCQ had poor to modest correlations with SF-36 (subscale) scores at one-year follow-up, ranging from 'MCS' (Spearman's r = -0.03, p > 0.05) to 'General health' (Spearman's r = -0.39, p < 0.0001). Total SCQ scores explained substantial variation for most SF-36 subscales, with R^2 values ranging from 0.10 ('Social function') to 0.25 ('Physical function') in multivariate (including age, sex, ethnicity, education level, and insurance status) linear regression models. Health care utilization: *Hospitalizations in previous year:* SCQ scores correlated fairly with hospitalizations in the previous 12 months (Spearman's r = 0.21, p < 0.01 for medical patients; Spearman's r = 0.37, p < 0.01 for surgical patients) *Prescription medication:* SCQ scores also correlated moderately with number of prescriptions (Spearman's r = 0.40 for medical patients; r = 0.55 for surgical patients) *Total hospital charges:* SCQ scores correlated poorly with total inpatient charges (Spearman's r=0.09 for medical patients; r = 0.10 for surgical patients) *Length of stay:* SCQ scores also correlated poorly with hospital length of stay (Spearman's r = 0.03 for medical patients; r = 0.14 for surgical patients)
Comparisons: **1. Medical record-derived CCI** **2. Michaud-Wolfe index** **Results:** SCQ had poor to fair correlations with the medical record-derived CCI (Spearman's r = 0.24, p < 0.05) and Michaud-Wolfe index (Spearman's r = 0.43, p < 0.05) mSCQ also had moderate correlations with CCI (Spearman's r = 0.36, p < 0.05) and Michaud-Wolfe index (Spearman's r = 0.57, p < 0.05)	**Comparisons:** **1. Medical record-derived CCI** **2. Michaud-Wolfe index** **Results:** Mortality: - PROM: *BASDAI (disease activity):* SCQ correlated moderately with disease activity (Spearman's r = 0.27, p < 0.05), while CCI correlated poorly (Spearman's r = 0.01). SCQ was significantly associated (OR = 1.73, 95% CI 1.25 – 2.40, p < 0.01) with low disease activity (BASDAI<4). *BASFI (physical function):* SCQ correlated moderately (Spearman's r = 0.43, p < 0.05) with physical function, but was significantly associated (β = 0.11, 95% CI 0.03 – 0.19, p = 0.01) with BASFI. *SF-36 (health-related quality of life):* SCQ correlated moderately (Spearman's r = -0.45, p < 0.05) with the PCS subscale, and was significantly associated (β = -0.72, p = 0.03) with PCS. CCI and Michaud-Wolfe indices were not significantly associated with BASFI and SF-36-PCS. Health care utilization: -

Table 2 | Detailed Overview of Included Studies. (continued)

	Items	Study Population (N)	Item-Level Reliability vs. Other Data Sources
Robinski[30]	German version (SCQ-G) of the SCQ instrument[27]	780 adult end-stage renal disease patients from 55 dialysis units. Data from the CORETH project.[31] **Characteristics:** Female 32.6% Mean age 63.2 years (± SD 15.1)	**Comparison: Medical record-derived CCI** **Results:** Kappa: 0.01 (peptic ulcer disease) – 0.84 (diabetes) Sensitivity: - Specificity: - PPV: - NPV: - Overall agreement between the SCQ and CCI ranged from 70 (heart disease) to 95% (kidney disease, liver disease). Positive agreement between both data sources ranged from 6% (peptic ulcer disease) to 97% (kidney disease). Negative agreement ranged from 78% (heart disease) to 97% (liver disease).
Bayliss[32]	Angina/ CAD Asthma Back pain (chronic) or sciatica Bronchitis (chronic)/ COPD Cancer (diagnosed within past 5 years) Cholesterol (elevated) Colon problem (e.g. diverticulitis/ irritable bowel) Congestive heart failure Diabetes Hard of hearing Hypertension Kidney disease Nerve condition Osteoarthritis Osteoporosis Overweight Poor circulation (e.g. peripheral vascular disease) Rheumatic disease (e.g. fibromyalgia or lupus) Rheumatoid arthritis Stomach problem (e.g. gastritis/ ulcer/ reflux) Stroke Thyroid disorder Vision problem	156 patients (≥ 65 years) from the HMO. **Characteristics:** Female 49.4% Mean age 75 years (67-94) Caucasian 91% College level or higher 59.6%	**Comparison: Medical records** **Results:** Kappa: - Sensitivity: 35 (kidney disease) – 100% (asthma) Specificity: 61 (hard of hearing) – 100% (kidney disease, cancer) PPV: - NPV: -

Overall Instrument Reliability vs. Other Data Sources	Evaluated Outcome(s)
Comparison: Medical record-derived CCI **Results:** Total SCQ-G score was moderately correlated to the medical record-derived CCI (Spearman's r = 0.27, p < 0.01).	**Comparison: Medical record-derived CCI** **Results:** Mortality: - PROM: *SF-12 (health-related quality of life):* Total SCQ-G score was moderately correlated with MSC (Spearman's r = -0.25, p < 0.01) and PSC (Spearman's r = -0.49, p < 0.01) subscales, while CCI correlated poorly with MSC (Spearman's r = 0.06, p > 0.05) and moderately with PSC (Spearman's r = -0.36, p < 0.01). Health care utilization: -
Comparison: Medical records **Results:** Sensitivity by respondent analysis: 14 – 100% (median 75%) Specificity by respondent analysis: 59 – 100% (median 91%)	**Comparisons:** **1. Medical records** **2. Medical record-derived CCI** **3. Rx-risk score (comorbidity measure including age, gender, health insurance benefit status and a category based on diagnoses from administrative pharmacy data)** **Results:** Mortality: - PROM: *SF-36 (health-related quality of life):* Self-reported number of conditions (Spearman's r = 0.477, p < 0.001) had a similar correlation compared to the medical-record CCI (r = 0.48, p < 0.001) but higher compared to the RxRisk score (0.17, p = 0.037). *SF-36 (physical functioning):* Self-reported conditions (r = -0.482, p < 0.001) had a stronger correlation vs. CCI (r = -0.41, p < 0.001) and the RxRisk score (r = -0.18, p = 0.035). *BRFSS (depression screening):* Self-reported conditions (r = -0.24, p = 0.003) had a stronger correlation compared to CCI (r = -0.12, p = 0.14) and the a RxRisk score (r = -0.05, p = 0.559). *GSE (self-efficacy):* Self-reported conditions (r = -0.305, p < 0.001) had a stronger correlation vs. CCI (r = -0.14, p = 0.096) and the RxRisk score (r = 0.10, p = 0.234). Health care utilization: -

Table 2 | Detailed Overview of Included Studies. (continued)

	Items	Study Population (N)	Item-Level Reliability vs. Other Data Sources
Poitras[33]	French modified version (DBMA-Fv) of the DBMA instrument[32]: 21 of the 23 original conditions were chosen Items "Kidney disease" and "Nerve condition" were excluded in this version Item "Depression/ anxiety" was added to this version	78 patients from 1 health center. **Characteristics:** Female 68% Mean age 47.4 years (± SD 15.9) College level or higher 70.5%	**Comparison: Medical records** **Results:** Kappa: - Sensitivity: 62.5 (angina/ CAD) – 90% (diabetes) - Specificity: 77.6 (overweight) – 98.6% (diabetes) PPV: 44.4 (overweight) – 92.9% (hypercholesterolemia) NPV: 88.7 (osteoarthritis) – 95.9% (asthma/ diabetes)
Wijers[34]	Spanish modified version of the DBMA instrument[32]: 21 out of the 23 original conditions were chosen Item "Liver disease" was excluded from this version due to low prevalence Items "UTI", "anxiety" and "memory-related disorders" were added to this version due to high due to high prevalence in older adults	707 community-dwelling adults (≥ 65 years). Data from the ELES-PS study.[35] **Characteristics:** Female 57% Mean age 74.2 years (± 6.6) College level or higher 17.3%	-
Simpson[36]	Angina pectoris Arthritis (OA/ RA) Cancer Congestive heart failure Diabetes mellitus Disc disease Hip fracture Lung disease Myocardial infarction Osteoporosis Parkinson's disease Peripheral arterial disease Spinal stenosis Stroke	1002 disabled women, aged ≥ 65 years, with MMSE ≥18. Data from the Women's Health and Aging Study I.[37] **Characteristics:** Age group 65-74 years 44.2% Caucasian 71.1%	Comparison: Disease-specific standardized algorithms (medical history, physical examination, medication review, medical record review, X-rays, physician questionnaire) **Results:** Kappa: 0.24 (peripheral arterial disease) – 0.96 (hip fracture) Sensitivity: 22 (spinal stenosis) – 98% (stroke) Specificity: 45 (arthritis) – 100% (hip fracture, Parkinson's disease, disc disease, spinal stenosis) PPV: 0.20 (peripheral arterial disease) – 1.0 (Parkinson's disease) NPV: 0.38 (arthritis) – 1.0 (hip fracture, Parkinson's disease, cancer, stroke)

Overall Instrument Reliability vs. Other Data Sources	Evaluated Outcome(s)
Comparison: Cumulative Illness Rating Scale **Results:** DBMA-Fv correlated moderately with the CIRS at baseline (r = 0.46, 95% CI 0.26 – 0.62, p < 0.01) and at two weeks follow-up (Spearman's r = 0.56, 95% CI 0.38 – 0.70, p < 0.01). **Comparison: Medical records** **Results:** Mean sensitivity of patient-reported conditions vs. medical record review at two weeks was 73.9% (± SD 8.4), while mean specificity was 92.2% (± SD 6.7).	-
-	**Comparison: Self-reported conditions** **Results:** Mortality: - PROM: *PWI (physical functioning & perceived health status):* DBMA significantly correlated stronger to physical functioning than self-reported number of conditions (Spearman's r = -0.56 vs. r = -0.51, p = 0.0035) *PWI (quality of life):* DBMA significantly correlated stronger to PWI in comparison to self-reported number of conditions (Spearman's r = -0.41 vs. r = -0.35, p = 0.0006) *CES-D (depression screening):* DBMA significantly correlated stronger to CES-D compared to self-reported number of conditions (Spearman's r = 0.41 vs. r = 0.35, p = 0.0043) Health care utilization: -
-	-

Table 2 | Detailed Overview of Included Studies. (continued)

	Items	Study Population (N)	Item-Level Reliability vs. Other Data Sources
Crabtree[38]	Angina Anxiety and depression Any other condition Arthritis/ Osteoporosis Breathlessness secondary to cardiovascular cause Breathlessness/ Wheeze (secondary to respiratory cause) Cerebrovascular disease Constipation Cough/ Sputum (secondary to COPD/ Asthma) Diabetes Diarrhea Epilepsy Hearing problems Pain Parkinson's disease Peripheral vascular disease Side effects from medications Skin disease Unsteadiness, falls and syncope Upper gastrointestinal symptoms Urinary problems Visual problems Walking and mobility	183 patients ≥ 65 years with confirmed age-related cataract (N=161) or from a geriatric day hospital (N=22). **Characteristics: -**	-
De-loyde[39]	Another cancer Chronic respiratory disease Depression Diabetes Heart disease Hypertension Kidney disease	756 patients with colorectal cancer from multiple hospitals. Data from the CONNECT[40] RCT. **Characteristics:** Female 56% Age group <70 years 54% College level or higher 24%	**Comparison: Clinician report** **Results:** Kappa: 0.22 (another cancer) – 0.58 (diabetes) Sensitivity: - Specificity: - PPV: - NPV: - **Comparison: Medical records** **Results:** Kappa: 0.34 (kidney disease) – 0.77 (diabetes) Sensitivity: - Specificity: - PPV: - NPV: -

Overall Instrument Reliability vs. Other Data Sources	Evaluated Outcome(s)
-	**Comparison:** - **Results:** Mortality: - PROM: *NEADL (activities of daily living):* CmSS correlated moderately to the NEADL (Spearman's r = 0.56, p < 0.01). *GHQ-28 (perceived health status):* CmSS correlated poorly to the GHQ-28 (Spearman's r = 0.48, p < 0.01) *HADS (anxiety and depression):* CmSS correlated moderately to the HADS (Spearman's r = 0.52, p < 0.01). Health care utilization: -
-	-

Table 2 | Detailed Overview of Included Studies. (continued)

	Items	Study Population (N)	Item-Level Reliability vs. Other Data Sources
Gad[41]	Amputation(s) Anemia or other blood problem(s) Asthma/ Other lung disease Back pain Blood clots or phlebitis Bowel problems Cancer Chronic skin condition Congestive heart failure Depression or anxiety Diabetes Excessive weight Hearing loss Heart attack High blood pressure High cholesterol Kidney or urinary problems Liver/ Gallbladder disease Lupus/ Other autoimmune disease Neuromuscular disease Osteoarthritis/ Degenerative arthritis Osteoporosis Paralysis Peripheral vascular disease Previous fracture(s) Recent unwanted weight loss Rheumatoid arthritis Sleep problems Stroke Thyroid problems Ulcer/ Stomach problems Visual problems	382 preoperative orthopedic patients (aged ≥ 65 years) before undergoing total knee or hip arthroplasty. **Characteristics:** Female 65% Mean age 74 years (± SD 6.1)	**Comparison: Medical records** **Results:** Kappa: 0.00 (osteoarthritis) – 0.76 (diabetes) Sensitivity: 9 (peripheral vascular disease) – 71% (hypertension) Specificity: 44 (osteoarthritis) – 99% (diabetes) PPV: - NPV: -

Overall Instrument Reliability vs. Other Data Sources	Evaluated Outcome(s)
-	-

Table 2 | Detailed Overview of Included Studies. (continued)

	Items	Study Population (N)	Item-Level Reliability vs. Other Data Sources
Hansen[42]	*For analysis (32/46 diagnosis groups):* Anemia Asthma/ COPD Atherosclerosis/ PAOD Cancers Cardiac arrhythmias Cardiac insufficiency Cardiac valve disorders Cerebral ischemia/ Chronic stroke Chronic cholecystitis/ Gallstones Chronic ischemic heart disease Chronic low back pain Diabetes mellitus Dizziness Gynecological problems Hemorrhoids Hypertension Hyperuricemia/ Gout Intestinal diverticulosis Joint arthrosis Lipid metabolism disorders Lower limb varicosis Migraine/ Chronic headache Neuropathies Osteoporosis Parkinson's disease Prostatic hyperplasia Psoriasis Renal insufficiency Rheumatoid arthritis/ Chronic polyarthritis Severe vision reduction Thyroid dysfunction Urinary tract calculi	3189 multi-morbid primary care patients. Data from the Multi-Care Cohort Study.[43] **Characteristics:** Female 59.3% Mean age 74.4 years (± SD 5.2) College level or higher 10.9%	**Comparison: Clinician report** **Results:** Kappa: 0.05 (gynecological problems) – 0.80 (diabetes) Sensitivity: - Specificity: - PPV: - NPV: -

Overall Instrument Reliability vs. Other Data Sources	Evaluated Outcome(s)
-	-

Table 2 | Detailed Overview of Included Studies. (continued)

	Items	Study Population (N)	Item-Level Reliability vs. Other Data Sources
Horton[44]	Anemia Anxiety Arthritis Bipolar disorder Breast cancer Cataracts Colon cancer Depression Diabetes Epilepsy Fibromyalgia Glaucoma Heart disease Hip replacement Hyperlipidemia Hypertension Inflammatory bowel disease Irritable bowel syndrome Kidney disease Knee replacement Liver disease Lung cancer Lung disease Migraine Osteoporosis Peptic ulcer disease Peripheral vascular disease Rectal cancer Rheumatoid arthritis Schizophrenia Sjogren's syndrome Skin cancer Systemic lupus erythematosus Thyroid Uveitis Vitamin-B12 deficiency	404 patients with multiple sclerosis from 2 centers. **Characteristics:** Female 76% Mean age 46.5 years (± SD 11.8) Caucasian 92% College level or higher 63.4% Relapsing-remitting MS 70.8%	**Comparison: Medical records** **Results:** Kappa: 0.19 (anemia) – 0.88 (diabetes) Sensitivity: 14 (kidney disease) – 100% (bipolar disorder, breast cancer, glaucoma, lung cancer, rheumatoid arthritis, schizophrenia, cataracts) Specificity: 87 (depression) – 100% (breast cancer, lung cancer) PPV: 0.07 (skin cancer) – 1.00 (breast cancer/ lung cancer) NPV: 0.84 (depression) – 1.00 (bipolar disorder, breast cancer, cataracts, glaucoma, lung cancer, rheumatoid arthritis, schizophrenia)
Iecovich[45]	Arthritis Cancer Cardiovascular accident Circulatory disease Diabetes Gastrointestinal disease Hypertension Myocardial infarction Osteoporosis Other heart diseases Renal problems Respiratory disease Thyroid disease	402 disabled older patients who used adult daycare centers. **Characteristics:** Female 74.8% Mean age 78 years (± SD 7.02) Asian/ African 62.6% College level or higher 10.1%	**Comparison: Medical records (including diagnostic ICD-9 codes)** **Results:** Kappa: 0.09 (circulatory disease) – 0.76 (diabetes) Sensitivity: 22.5 (cancer) - 79.1% (diabetes) - Specificity: 73.5 (renal) – 98% (cancer) PPV: 0.36 (circulatory disease) – 0.92 (hypertension) NPV: 0.42 (hypertension) – 0.87 (thyroid disorder)

Overall Instrument Reliability vs. Other Data Sources	Evaluated Outcome(s)
Comparison: Medical records **Results:** Agreement between self-reports and medical records was κ = 0.56 (95% CI 0.48 – 0.64) for the presence of any physical comorbidity, and κ = 0.57 (95% CI 0.48 – 0.65) for mental comorbidities. For this analysis, the questionnaire was divided into physical vs. mental comorbid conditions, and thereafter dichotomized in 0 vs. >0 comorbidities.	-
Comparison: Medical records (including diagnostic ICD-9 codes) **Results:** Self-reports correlated fairly with the EHR (r = 0.45, p < 0.001).	-

Table 2 | Detailed Overview of Included Studies. (continued)

	Items	Study Population (N)	Item-Level Reliability vs. Other Data Sources
Klabunde[5]	Angina Arthritis or rheumatism Chronic Lung Disease/ Bronchitis/ Emphysema Cirrhosis/ Liver disease Congestive heart failure Depression or anxiety Diabetes Hypertension IBD/ Colitis/ Crohn's disease Myocardial infarction Stroke/ Brain hemorrhage Stomach ulcers with bleeding	3095 prostate cancer survivors. Data from the PCOS study.[46] **Characteristics:** Age group >65 years 64% Caucasian 78% College level or higher 60%	-
Boissonnault[47]	Anemia Ankylosing spondylitis Arterial blockage of legs Asthma Cancer Chemical dependency Deep venous thrombosis Degenerative osteoarthritis or wear-and-tear arthritis Depression Diabetes (diagnosed after age 18 years) Diabetes (diagnosed before age 18 years) Emphysema Endometriosis Epilepsy/ Seizures Gout Headaches (>1 per week) Heart attack Heart valve problems Hepatitis Hypertension Hyperthyroid Hypothyroid Infections Multiple sclerosis Osteoporosis Other illnesses (please list) Rheumatoid arthritis Stomach/ Duodenal ulcers Stroke Tuberculosis Urinary incontinence Questionnaire contains 91 items divided into 8 sections (comorbidities, surgeries, medication, substance use and demographic characteristics)	100 preoperative orthopedic surgery patients at 1 hospital. **Characteristics:** Female 54% Mean age 46.9 years (± SD 16.7) College level or higher 64%	**Comparison: NP/PA responses to identical questionnaire following medical record review and/or patient interview** **Results:** Kappa: 0.15 (other illnesses) – 1.00; mean $\kappa = 0.69$ Sensitivity: - Specificity: - PPV: - NPV: -

Overall Instrument Reliability vs. Other Data Sources	Evaluated Outcome(s)
-	-
Comparison: NP/PA responses to identical questionnaire following medical record review and/or patient interview **Results:** Mean percentage agreement across all questionnaire items between self-report and NP/PA report was 96%.	-

Table 2 | Detailed Overview of Included Studies. (continued)

	Items	Study Population (N)	Item-Level Reliability vs. Other Data Sources
Fan[48]	Angina Arthritis CABG/ PTCA Cancer Congestive heart failure Coronary artery disease Depression Diabetes Drug abuse Enlarged prostate Heartburn HIV Hypertension Liver disease Lung disease Osteoporosis Pneumonia Post-Traumatic Stress Disorder Prior myocardial infarction Renal insufficiency Seizure Stroke Thyroid disease Ulcer disease	Development sample: 5469 patients from 7 VA medical centers. Data from the ACQUIP study.[49] **Characteristics:** Female 2.5% Mean age 67.8 years (± SD 0.1) Caucasian 83.4% College level or higher 68.7% Validation sample: 5478 patients from 7 VA medical centers **Characteristics:** Female 2.7% Mean age 67.8 years (± SD 0.1) Caucasian 83.3% College level or higher 68%	-
Lorem[50]	*All respondents:* Angina Asthma Atopic eczema Cancer survivor Cerebrovascular stroke Chronic bronchitis Diabetes Duodenal ulcer Epilepsy Fibromyalgia Food allergies Hand eczema Hypersensitivity Kidney stone Liver disease Migraine Myocardial infarction Osteoporosis Pollen allergies Psoriasis Thyroid Ventricular ulcer *For patients >70 years, added:* Arthritis Cataract Glaucoma Parkinson's disease Rheumatoid arthritis Urinary incontinence	Reference population: 26684 patients sampled from Tromsø study (1994/ 1995).[51] **Characteristics:** Female 52.6% Age group <50 years 61.7% Validation population: 804 patients sampled from Tromsø study and FHI panel (2001/2002). **Characteristics:** Female 55% Ages 30-79 years	-

Overall Instrument Reliability vs. Other Data Sources	Evaluated Outcome(s)
-	**Comparison:** - **Results:** Mortality: *All-cause mortality:* SIC had a moderate discriminative ability (AUC = 0.71) of SIC in predicting mortality at 2 years follow-up. A combined model, containing SIC and SF-36 as predictors, had an AUC = 0.74. PROM: - Health care utilization: *Re-hospitalizations:* Discriminative ability of SIC was less able in predicting 2-year re-hospitalizations (AUC = 0.61), which slightly increased when SF-36 was added to the model (AUC = 0.64).
Comparison: Medical record-derived CCI **Results:** HII correlated more strongly with SRH vs. CCI (Spearman's $r = -0.360$, $p < 0.001$ vs. $r = -0.250$, $p < 0.001$). After excluding all patients with HII=0, the correlation between HII and SRH strengthened ($r = -0.421$, $p < 0.001$) while it weakened between CCI and SRH ($r = -0.141$, $p < 0.001$).	**Comparison:** - **Results:** Mortality: - PROM: *SRH:* In an ordinal logistic regression model (containing age, gender, mental health symptoms and the HII), HII had a negative effect ($\beta = -0.249$, $p < 0.001$) after adjustment for the other variables. Health care utilization: -

Table 2 | Detailed Overview of Included Studies. (continued)

	Items	Study Population (N)	Item-Level Reliability vs. Other Data Sources
Lucke[52]	*For analysis:* Asthma Cardiovascular disorder (combined) Coronary heart disease Diabetes mellitus Dyslipidemia GI Hypertension Hyperuricemia Mental disorders Osteoporosis Original questionnaire comprised of 51 (combined) diseases, as well as free text.	2653 patients with COPD or chronic bronchitis. Data from the COSYCONET study.[53] **Characteristics:** Male 59.4% Mean age 65 years (± SD 8.6) Mean BMI 27 (± SD 5.4) GOLD ≤ 2 57.7%	**Comparison: ATC-codes for disease-specific medication** **Results:** Concordance between self-reported comorbidities and ATC codes for disease-specific medication varied from 1.3% (asthma) to 51.8% (combined CVD).
Md Yusof[54]	18 sections: 12 physical systems 6 mental health state an additional free text option for medication prescription.	113 community-dwelling patients from 1 research center. **Characteristics:** Female 56.6% Mean age 75.3 years (± SD 5.19)	-
Merkin[55]	Angioplasty or CABG Cancer Cerebrovascular disease Congestive Heart Failure COPD Diabetes Hypertension Myocardial infarction	965 patients with ESRD from 81 dialysis clinics. Data from the CHOICE study.[56] **Characteristics:** Female 46% Mean age 58 years Caucasian 67% College level or higher 36%	**Comparison: Medical record-derived ICED** **Results:** Kappa 0.19 (hypertension) – 0.93 (diabetes) Sensitivity: 18 (COPD) – 96% (diabetes) Specificity: 76 (hypertension) – 98% (diabetes) PPV: - NPV: - **Comparison: Clinician report** **Results:** Kappa 0.19 (hypertension) – 0.81 (diabetes) Sensitivity: - Specificity: - PPV: - NPV: -

Overall Instrument Reliability vs. Other Data Sources	Evaluated Outcome(s)
Comparison: Matched ICD10-codes for diseases and non-specific medications **Results:** About 51.5% of self-reported comorbidities were confirmed after comparing them to matched ICD10-codes.	-
Comparison: Clinician report **Results:** CMI correlated significantly with the GP-data ($r = 0.8$, $p < 0.001$).	**Comparison:** - **Results:** Mortality: *Survival:* In a Cox proportional hazards model containing 5 continuous predictors (age, weighted CCI, combined condition and age-related CCI score, total score physical sections of CMI, total score mental sections of CMI, and count of medication prescriptions), none of these predictors significantly contributed to predicting mortality. PROM: - Health care utilization: -
-	-

Table 2 | Detailed Overview of Included Studies. (continued)

	Items	Study Population (N)	Item-Level Reliability vs. Other Data Sources
Mukerji[57]	Arthritis Diabetes Heart disease Lung disease Other cancers (apart from index tumor) Psychiatric problems Stroke	458 patients with newly diagnosed head-and-neck cancer from 3 hospitals. **Characteristics:** Female 23.6% Caucasian 86% College level or higher 49.8%	**Comparison: Medical record-derived ACE-27 index** **Results:** Kappa: 0.11 (arthritis) – 0.89 (diabetes) Sensitivity: - Specificity: - PPV: - NPV: - Over-reporting by self-report occurred more often (21.2%) compared to under-reporting (13.5%), with medical records as gold standard.
Paleri[58]	*9 section headers:* Heart & blood vessels Alcohol consumption Brain and nerves Cancer Diabetes Joints and muscles Kidney Liver, stomach and pancreas Lungs	20 patients with head-and-neck cancer. **Characteristics: -**	-
Pati[59]	Acid peptic disease Arthritis Asthma Cancer Chronic back ache Chronic kidney disease Chronic liver disease (alcohol) Deafness Dementia Diabetes Epilepsy Filariasis Heart disease Hypertension Stroke Thyroid Tuberculosis Visual difficulty	103 patients from 4 primary care practices. **Characteristics:** Female 45% Mean age 45 years (± SD 5.32)	**Comparison: Clinician's prescription-derived data** **Results:** Kappa: 0.58 (hearing problem) – 1.00 (tuberculosis); 16/18 conditions were evaluated. Sensitivity: - Specificity: - PPV: - NPV: -

Overall Instrument Reliability vs. Other Data Sources	Evaluated Outcome(s)
Comparison: Medical record-derived ACE-27 index **Results:** Kappa: 0.50 (0.44 – 0.57)	-
Comparison: Medical record-derived ACE-27 grade **Results:** Kappa: 0.92 (95% CI 0.82 – 1.0)	-
-	-

Table 2 | Detailed Overview of Included Studies. (continued)

	Items	Study Population (N)	Item-Level Reliability vs. Other Data Sources
Lee[60]	Alcohol use Arthritis BMI < 25 Chronic lung disease Current tobacco use Diabetes mellitus Heart failure History of falls History of pain Hypertension Incontinence Memory-related disease Non-skin cancers Other heart problems Psychiatric disease Stroke Visual or hearing impairment	Older (>50 years) community-dwelling patients. Data from the Health and Retirement Study.[61, 62] Development cohort: 11701 patients **Characteristics:** Female 57% Mean age 67 years (± SD 10) Caucasian 81% High school level or higher 75% Validation Cohort: 8009 patients **Characteristics:** Female 56% Mean age 67 years (± SD 10) Caucasian 71% High school level or higher 66%	-
Voaklander[63]	Asthma Cancer Chronic back pain Circulatory problems Diabetes Digestive problems Emphysema Epilepsy Eye problems Hay fever/Allergies Heart disease Hypertension Kidney disease Liver disease Other Stomach ulcers Stroke Thyroid problems	518 patients receiving major joint arthroplasty at 2 acute care facilities (1995-1997). 283 patients receiving TKA. **Characteristics:** Female 59% Mean age 69 years (± SD 9) 235 patients receiving THA. **Characteristics:** Female 60% Mean age 67.1 years (± SD 12)	-

Overall Instrument Reliability vs. Other Data Sources	Evaluated Outcome(s)
-	**Results:** - Mortality: *All-cause mortality:* a final model, predicting 4-year mortality, included 12 variables (6 comorbid conditions, sex, age, and 4 functional status measures). Discriminative ability of the model was determined in the development (ROC = 0.84) and validation cohort (ROC = 0.82). - PROM: - - Health care utilization: -
Comparisons: **1. Medical record-derived CCI** **2. Administrative data-derived CCI** **3. ICD-9 Sum of comorbidities (available in the EHR)** **Results:** SRC correlated significantly with the medical record-derived CCI (Spearman's r = 0.40, p < 0.01), administrative data-derived CCI (Spearman's r = 0.32, p < 0.01) and ICD-9 Sum (Spearman's r = 0.39, p < 0.01).	**Comparisons:** **1. Medical record-derived CCI** **2. Administrative data-derived CCI** **3. ICD-9 Sum of comorbidities** **Results:** Mortality: - PROM: *SF-36 & WOMAC (health-related quality of life):* SRC explained similar amounts in HRQoL domains compared to the ICD-9 Sum, medical record-derived CCI and administrative data-derived CCI. Health care utilization: *Hospital stay:* SRC explained the variance in acute length of stay less better compared to the other comorbidity measures. *ER visits:* SRC was slightly better in predicting emergency department visits than the other measures (2% variance).

Table 2 | Detailed Overview of Included Studies. (continued)

	Items	Study Population (N)	Item-Level Reliability vs. Other Data Sources
Selim[64]	*Physical conditions:* Angina pectoris Cancer Cataract Chronic low back pain Chronic lung disease Congestive heart failure Diabetes Diverticulitis Enlarged prostate Gallbladder disease Gout Heart attack Hepatitis High blood pressure Inflammatory bowel disease Irregular heartbeat Osteoarthritis Osteoporosis Peptic ulcer Peripheral vascular disease Phlebitis Prostatitis Renal failure Rheumatoid arthritis Seizure Skin cancer Stroke Thyroid disease Transient ischemic attacks Urinary tract infection *Mental conditions:* Alcohol use Anxiety Bipolar Depression Posttraumatic stress disorder Schizophrenia	2425 patients who previously received ambulatory care. Data from Veterans Health Study. **Characteristics:** Female 0% Mean age 64 years (± SD 12.7) High school level or higher 41%	-
Vigen[65]	Arthritis Cardiovascular disease Diabetes Gall bladder disease High cholesterol Hypertension Intestinal polyps Irritable bowel syndrome Osteoporosis Other cancers Thyroid disorders	1936 breast cancer patients. Data from the SFBCS & LACE studies.[66, 67] **Characteristics:** Mean age 60.4 years (± SD 11.0) Caucasian 71% College level or higher 67.8%	**Comparison: Medical records (including diagnostic ICD-9 codes)** **Results:** Kappa: 0.50 (other heart diseases) – 0.87 (diabetes) Sensitivity: 48 (other heart diseases) – 90.5% (myocardial infarction) Specificity: 96 (other heart diseases) – 99.5% (diabetes) PPV: - NPV: -

Overall Instrument Reliability vs. Other Data Sources	Evaluated Outcome(s)
-	**Comparison:** - **Results:** Mortality: *Survival:* In a Cox proportional hazards model, the mortality risk was 14% for every increment in physical CI which decreased to 9% ($p < 0.05$) after adjustment for sociodemographic and disability rating. PROM: *SF-36 (health-related quality of life):* Physical CI had the strongest correlations (Pearson's r ≥ 0.29) with SF-36-PCS, while the mental CI correlated better with SF-36-MCS (Pearson's r ≥ 0.30). Both physical and mental CI's were significantly correlated with all SF-36 scales. Combined physical/mental CI also correlated significantly with all HRQoL scales (Pearson's r range -0.29 – -0.45). Health care utilization: *Hospital visits:* Both physical and mental CI's had significant coefficients and explained 5% of the variance in total outpatient clinic visits, which increased to 7% after adjustment for sociodemographic and disability rating.
-	**Comparison: Medical records (including diagnostic ICD-9 codes)** **Results:** Mortality: *All-cause mortality:* No significant differences in HR's for covariates between comorbidity models with self-report vs. medical records as data source (diabetes HR 1.65 vs. 1.44; hypertension 1.22 vs. 1.55; myocardial infarction 1.40 vs. 1.73; other heart diseases 1.07 vs. 1.51) were observed. PROM: - Health care utilization: -

Table 2: Detailed Overview of Included Studies. (continued)

Note. – = unknown or not described in study, β = βeta

ACE-27, Adult Comorbidity Evaluation –27; *ADL*, Activities of Daily Living Scale; *ACQUIP*, Ambulatory Care Quality Improvement Project; *AIDS*, acquired immunodeficiency syndrome; *ASI*, Activity Status Index; *ATC*, Anatomical Therapeutic Chemical code; *AUC*, area under the curve; *BASDAI*, Bath Ankylosing Spondylitis Disease Activity Index; *BASFI*, Bath Ankylosing Spondylitis Functional Index; *BMI*, body mass index; *BRFSS*, Behavior Risk Factor Surveillance System; *CABG*, coronary artery bypass graft; *CAD*, coronary artery disease; *CCI*, Charlson Comorbidity Index; *CES-D*, Center for Epidemiologic Studies Depression Scale; *CHF*, congestive heart failure; *CHOICE*, Choices for Healthy Outcomes in Caring for End-stage renal disease study; *CI*, confidence interval; *CIRS*, Cumulative Illness Rating Scale; *CMI*, Cornell Medical Index; *CmSS*, Comorbidity Symptom scale; *CONNECT*, Centralized Nurse-Led Telephone-Based Care Coordination to Improve Outcomes After Surgical Resection for Colorectal Cancer; *COPD*, chronic obstructive pulmonary disease; *CORETH*, the Choice of Renal Replacement Therapy project; *COSYCONET*, COPD and Systemic Consequences – Comorbidities Network Cohort Study; *CTD*, connective tissue disease; *CVA*, cerebrovascular accident; *CVD*, cardiovascular disorder; *DBMA*, Disease Burden Morbidity Assessment; *DMBA-Fv*, French version of the Disease Burden Morbidity Assessment; *EHR*, electronic health record; *ELES-PS*, Aging in Spain Longitudinal study, Pilot Survey; *EQ-5D*, EuroQol-5 Dimension; *ER*, emergency room; *ESRD*, end-stage renal disease; *FHI*, Norwegian Institute of Public Health; *GHQ-28*, General Health Questionnaire –28; *GOLD*, Global Initiative for Chronic Obstructive Lung Disease; *GP*, general practitioner; *GSE*, general self-efficacy; *HADS*, Hospital Anxiety and Depression Scale; *HII*, Health Impact Index; *HIV*, human immunodeficiency virus; *HMO*, Health Maintenance Organization; *HR*, hazard ratio; *HRQol*, health-related quality-of-life; IBD, inflammatory bowel disease; *ICC*, intra-class correlation coefficient; *ICD-9*, International Classification of Diseases version 9; *ICD-10*, International Classification of Diseases version 10; *ICED*, Index of Coexistent Disease; *LACE*, Life After Cancer Epidemiology study; *MCS*, mental component summary; *MMSE*, Mini Mental Status Examination; *MS*, multiple sclerosis; *MSC*, mental sum scale; *mSCQ*, modified version of the Self-Administered Comorbidity Questionnaire; *NEADL*, Nottingham Extended Activities of Daily Living; *NP*, nurse practitioner; *NPV*, negative predictive value; *OA*, osteoarthritis; *OASIS*, Outcome in Ankylosing Spondylitis International Study; *OR*, odds ratio; *PA*, physician assistant; *PAOD*, peripheral arterial occlusive disease; *PCOS*, Prostate Cancer Outcomes Study; *PID*, pelvic inflammatory disease; *PPV*, positive predictive value; *PROM*, patient-reported outcome measure; *PSC*, Physical Sum Scale; *PTCA*, percutaneous transluminal coronary angioplasty; *PWI*, Personal Wellbeing Index; Spearman's *r*, Spearman correlation coefficient; *RA*, rheumatoid arthritis; *RCT*, randomized controlled trial; *ROC*, receiver-operating characteristic; *Rx*, prescription; *SCQ*, Self-Administered Comorbidity Questionnaire; *SCQ-G*, German version of the Self-Administered Comorbidity Questionnaire; *SD*, standard deviation; *SF-36-MCS*, Short-Form 36 Mental Component Summary scale; *SF-36-PCS*, Short-Form 36 Physical Component Summary scale; *SFBCS*, San Francisco Bay Area Cancer Study; *SIC*, Seattle Index of Co-morbidity; *SRC*, Self-Reported Co-Morbidity; *SRH*, Self-Reported Health; *THA*, total hip arthroplasty; *TIA*, transient ischemic attack; *TKA*, total knee arthroplasty; *UTI*, urinary tract infection; *VA*, Veterans Affairs; *WOMAC*, Western Ontario McMaster Osteoarthritis Index.

Presence of Specific Conditions and Related Assessments

Conditions that were most commonly included in the patient-reported morbidity instruments were diabetes, hypertension, myocardial infarction, and stroke (see Table 2). The question regarding the presence or absence of specific comorbidities was presented in multiple ways, including:

- "Do you have or have you ever had…?"[32, 34, 41, 50, 60, 65]
- "Has a doctor ever told you that you have…?"[5, 44, 45, 48, 55, 59, 60]
- "Do have you any of the following problems?"[27, 28, 30, 45]

Most instruments had close-ended response alternatives for each condition listed. Some questionnaires had an additional free-text item for patients to report additional comorbidities that were not listed in the instrument.[27, 33, 47, 52, 63] Several instruments included additional questions regarding the severity of the conditions, such as, "Does it limit your activities?"[5, 27, 28, 30, 32, 34] or regarding active treatment, for example, "Do you receive treatment for it?"[5, 27, 28, 30, 44]

Instrument Administration, Length, Duration, and Responsiveness

Nineteen studies had self-administered questionnaires (either in mailed/ written or electronic form),[5, 25-28, 30, 32, 33, 34, 36, 39, 41, 44, 47, 48, 50, 52, 54, 57] whereas 8 studies had questionnaires that were administered verbally by a clinical/ research associate (either face-to-face or by phone)[12, 22, 38, 42, 45, 60, 64] and 4 studies reported both administration methods.[9, 11, 65, 24] Of the associate-administered surveys, it was not clear whether clarifying questions were allowed or used in nearly all of the studies. Five studies[9, 11, 24, 57, 65] had comorbidity questionnaires that could either be self-administered or administered by an interviewer (e.g. associate or other proxy) if needed (e.g. for patient illiteracy).

The length of the instruments varied from 4 items[65] to 195 items[54] (divided over multiple physical and mental sections). Nine studies mentioned the duration to complete the questionnaire, which ranged from 1 minute[26] to 45 minutes.[45] Response rates were provided in 9 studies ranging from 28%[32] to 99%.[30]

Reliability and Concordance with Other Data Sources

Test-retest reliability was described in 7 studies[5, 11, 27, 33, 38, 47, 59] (data not shown), mostly measured by the intra-class correlation or Spearman correlation coefficients. The amount of patients on which it was tested ranged from 26[11, 27] to 103.[2] The interval period between both measurements varied from 24 hours[11, 27, 47] to 4 weeks.[38] The overall Spearman correlation coefficients for patient-reported comorbidity questionnaires ranged from 0.73[11] (moderate reliability) to 0.87 (strong

reliability),[38] whereas the intra-class correlation coefficients ranged from 0.86[33] to 0.97.[59]

Whole instrument and item-level concordance of patient-reported morbidity scores with information from other data sources, either medical records or medical record-derived comorbidity indices, were most frequently assessed (see Table 2). Spearman correlation coefficients for the relationship between patient-reported morbidity scores and composite scores from other data sources ranged from r = 0.24 (14 conditions)[28] to r = 0.70 (18 conditions).[11] In studies measuring Kappa coefficients, κ values were notably higher for agreement with medical records (κ range: 0.56-0.69)[44, 47] as opposed to agreement with medical record-derived morbidity indices (κ range: 0.37-0.50).[25, 57] Administrative data were also used as a comparative data source in a number of studies,[9, 26, 45, 52, 63] which generally demonstrated poor agreement.[9, 45, 63]

Item-level (single condition) was the most commonly reported form of concordance assessment, mostly measured against medical records or derived morbidity indices. A striking observation was that diabetes, as a comorbidity questionnaire item, had the highest Kappa value across included studies.[24, 26, 30, 39, 41, 42, 44, 45, 55, 57, 65] Most included studies had a substantially wide κ value range,[22, 24, 25, 26, 27, 28, 30, 36, 41, 42, 44, 45, 55, 57] in general from 0.66[27] to 0.86.[28]

Association with Health and Healthcare Outcomes

A number of studies assessed the association between patient-reported instrument scores and mortality, patient-reported outcome measures, and healthcare utilization. Because none of the included studies evaluated their instrument against all 3 outcomes, we provided some examples in this paragraph to demonstrate the directionality of the associations.

Some studies assessed the relationship between patient-reported instrument scores and mortality or survival.[25, 26, 48, 54, 60, 64, 65] Habbous et al[25] demonstrated a significant relation between the patient-reported Charlson Comorbidity Index (CCI) and overall survival, with the presence of at least 2 comorbidities being associated with worse survival (hazard ratio [HR] = 1.62, P=.003). Nevertheless, this relation was stronger for the nonpatient-reported, medical record-derived CCI (HR = 2.60). Only a few studies developed prediction models for all-cause mortality with patient-reported comorbidity instruments, either in combination with other predictors such as demographic variables[60] or by themselves.[26, 48] Fan et al[48] developed a prediction model with the Seattle Index of Comorbidity (SIC) and estimated

an area under the curve (AUC) = 0.71 for all-cause mortality at 2 years follow-up, whereas Lee et al[60] estimated an AUC = 0.82 of a different model (including sex, age, 6 comorbidities, and 4 functional measures) for all-cause mortality at 4 years follow-up.

The most common patient-reported outcome measure evaluated with patient-reported morbidity instruments was the 36-item Short Form Health Survey (SF-36).[24, 27, 28, 32, 63, 64] One study reported a higher patient-reported CCI score being negatively associated with SF-36 scores,[24] and another reported the number of self-reported conditions in the DBMA instrument being significantly correlated to the SF-36 similar to the medical record-derived CCI.[32] Selim et al[64] also demonstrated a negative association between high CI scores and SF-36 scores. Because of heterogeneous reports on the association and correlation units between both measures in the included studies, there was no clear consistency observed in the direction of the association or correlation.

Several studies also analyzed the relationship between patient-reported comorbidity instruments and healthcare utilization outcomes (e.g. [re]hospitalizations, emergency room visits, medical costs).[9, 11, 26, 27, 63] Susser et al[9] estimated an AUC of 0.68 and AUC of 0.67 in predicting the number of hospital days and emergency room visits, respectively. Katz et al[11] estimated weak correlations between the patient-reported comorbidity CCI and healthcare utilization outcomes, and Sangha et al[27] reported similar weak associations with the SCQ.

Impact of Demographic Factors on Survey Validity and Reliability

Thirteen studies reported associations between certain patient characteristics and concordance of comorbidity questionnaires with other sources of comorbidity data. Most studies reported higher age being significantly associated with lower agreement between patient-reported and medical-record derived comorbidity data.[22, 25, 36, 39, 42, 45, 47, 55, 57] Nevertheless, Katz et al,[11] Vigen et al,[65] and Horton et al[44] did not observe a significant association. In terms of reliability, Klabunde et al[5] found a significant association between age ≥ 65 years and

inconsistent response patterns between baseline and subsequent surveys. Higher concordance between patient-reported and other comorbidity data sources was also associated with higher education levels.[11, 39, 42, 45] This association was also observed for higher socioeconomic status.[5] In contrast, Vigen et al[65] only reported this association for myocardial infarction. Simpson et al[36] reported that education level did not impact reliability, with Kappa values remaining unaltered after

adjustment. Some studies[42, 45, 55] also reported concordance being significantly influenced by sex, whereas others[44] did not observe a significant observation.

DISCUSSION

Risk adjustment for benchmarking of healthcare outcomes across multiple hospitals is dependent on accurate reporting of case-mix factors such as patient morbidity (ie, comorbid conditions). Clinicians have increasingly considered patient-reported comorbidity instruments as a potential alternative to the laborious review of medical and administrative records or as a method to standardize the collection of data for important conditions across hospitals. Previously published research has looked at multi-morbidity measures in primary care or community population settings[68] but, to our knowledge, this is the first systematic literature review focusing on the development or subsequent validation of general patient-reported morbidity instruments (either indices or ad hoc lists of conditions).

Ten original patient-reported comorbidity instruments were found, as well as additional variations on these original studies. The most frequently cited instruments were the SR-CCI and the SCQ, with the SR-CCI demonstrating stronger item-level reliability, overall reliability, and overall agreement, but similar correlations with healthcare utilization parameters. The number of items varied substantially from instrument to instrument. Most studies evaluated the accuracy of self-reported comorbid conditions for individual items compared with another data source, most commonly medical records review. Agreement with the medical record, which was generally used as the gold standard, varied substantially based on the comorbid condition listed within all of the reviewed survey instruments. The kappa values regarding item-level validity were generally not used to eliminate questions with low reliability, leading to large intra-instrument variability. This variation by comorbid condition was thought by the authors to be the result of accurate medical record data with missed diagnoses by the patient and accurate reporting by the patient with missed data in the medical record. The authors postulated that disease items with low reliability included diseases that are "resolved" (in the past),65 those that are controlled with treatment (e.g. hypertension),[65] those without symptoms,[36] those with complex diagnostic criteria or ambiguous disease categories (e.g. heart diseases such as atherosclerosis or heart failure),[24, 65] and those with confusing or overlapping names (e.g. arthritis vs osteoarthritis vs rheumatoid arthritis).[24, 36, 44, 57] Violán et al[69] demonstrated similar results in a cross-sectional study comparing morbidity prevalence between electronic health records and health surveys;

self-reported morbidity prevalence was higher among younger patients and for symptomatic conditions. Diseases with clear definitions (e.g. diabetes) and that required ongoing treatment had higher agreement with other data sources and were most accurately reported by patients.[25, 36, 45] Even in those cases, there may have been disagreement, such as if people with pre-diabetes classified themselves as diabetic or people with non–insulin-dependent diabetes considered themselves not diabetic. Agreement between patient-reported morbidity instruments and administrative data was usually poor, with the limitation generally listed in some studies that administrative data may underreport the presence of comorbid conditions ("undercoding") more than the medical record.

This systematic review highlights the lack of information on the predictive validity of comorbidity data; a subset of included studies examined the predictive capabilities of the patient-reported comorbidity questionnaires/ indices for outcomes such as mortality, patient-reported outcome measures (such as functional status or general health-related quality of life), and healthcare utilization and costs. The patterns of correlations between patient-reported comorbidity data and mortality were as hypothesized: higher patient-reported comorbidity scores were associated with poorer overall survival.[25] The predictive ability of patient-reported comorbidity indices were moderate to good (AUCs > 0.70) for all-cause mortality, regardless of whether only comorbidities were used or other variables were added to the model.[26, 48, 60] Unsurprisingly, higher patient-reported comorbidity scores were significantly correlated and associated with lower health-related quality of life scores (as measured by the SF-36).[24, 27, 28, 32, 63, 64] Patient-reported comorbidity scores had low positive (Spearman's $r < 0.50$) correlations with many healthcare utilization measures (e.g. hospitalizations, length of hospital stay, prescribed medications)[11, 27] and had poor discriminative ability (AUCs between 0.60-0.70) for these measures as well.[9, 48] Because comorbidity measures can influence outcomes and interpretation when comparing treatment strategies or hospitals, it is essential that a comorbidity measure is validated for the population and outcome of interest.

Gold standard survey instrument development generally begins with a systematic exploration of potential topics to include in survey questions, followed by the development of a large number of questions, which overlap and then reduce to a smaller number of higher performing questions based on field testing.[70, 71] For example, a question on the presence or absence of myocardial infarction could be asked several ways, and the question that proves most valid and reliable in testing is retained and the other forms of the question are discarded. Although 7 stud-

ies mentioned that their questionnaires were rephrased to optimize clarity and comprehension,[27, 32, 33, 38, 44, 47, 59] there were no studies that started intentionally with a larger question bank that was then reduced down to reliable and valid questions.

This study should be interpreted in the context of its limitations. Although an extensive search in 5 electronic database was conducted with the help of an experienced librarian, the possibility of missing studies cannot be excluded. Comorbidity data is collected in a large proportion of medical studies, and therefore inclusion and exclusion criteria had to be defined to allow for reasonable identification of self-reported comorbidity instruments. The goal was to identify studies that primarily focused on instrument development or validation. It is possible that comorbidity questionnaires that were used but did not have any description of their development, validity, or reliability in the abstract or title could have been missed in the present search strategy, such as instruments in the gray literature. Additionally, studies that may have subsequently used information from a comorbidity instrument to predict an outcome in a general study without mention of the instrument itself in the abstract could have been missed with this search strategy. The presence of 2 independent reviewers to conduct the study selection, data extraction, and overall interpretation added to the accuracy of the review process.

This systematic review was inspired by international efforts (e.g. the International Consortium for Health Outcomes Measurement [ICHOM],[3] the United States Agency for Healthcare Research and Quality (AHRQ), [72] and the Registry of Patient Registries project[73]) to harmonize collection and format of outcome measures.

These outcome measures can include suggested outcome domains, measurement tools, and predictor variables for risk adjustment for a given medical condition. Among these risk-adjustment factors are morbidity variables, which can be collected and analyzed using various methods. Minimal standards for consistent capture of morbidity data are essential for fair benchmarking for VBHC or public reporting or outcomes. Patient-reported morbidity instruments can be used internationally and among hospitals with no or different electronic medical records and administrative data structure. Additionally, these instruments could be used to bolster the medical record. Future research should focus on the capture of complete morbidity data for the purposes of more robust risk adjustment, a key component of fair benchmarking for VBHC.

CONCLUSIONS

In summary, this systematic review found ten self-reported morbidity instruments, with the Patient-Reported Charlson Comorbidity Index[11] and the Self-Administered Comorbidity Questionnaire[27] being the most frequently cited instruments. Within each included instrument, there was significant variability in the reliability of patient-reported comorbidities based on the comorbid condition. Further research is needed to determine whether patient-reported comorbidity data should be used to bolster medical records data or serve as a stand-alone entity for risk adjustment of observational outcomes data.

ACKNOWLEDGMENTS

- The authors would like to acknowledge Wichor M. Bramer (Biomedical Information Specialist, Medical Library, Erasmus MC) for his support in the literature search.
- Authors A. Oemrawsingh and J.A. Hazelzet were supported by a grant from the Federation of Dutch University Medical Centers (NFU).
- R.W. Bergmark received grants from the Gliklich Healthcare Innovation Scholar Fund/ Massachusetts Eye and Ear Infirmary and from the American Board of Medical Specialties during the development of this paper.

CONFLICTS OF INTEREST

- Author N. Swami was a paid employee at the International Consortium for Health Outcomes Measurement (ICHOM) during the development of this paper.
- Author A.L. Pusic reports personal fees from Patient Reported Outcome Measures (Q-PROMS) outside the submitted work.

REFERENCES

1. Fortin M, Soubhi H, Hudon C, Bayliss EA, van den Akker M. Multimorbidity's many challenges. *BMJ*. 2007;334(7602):1016–1017.
2. Valderas JM, Starfield B, Sibbald B, Salisbury C, Roland M. Defining comorbidity: implications for understanding health and health services. *Ann Fam Med*. 2009;7(4):357–363.
3. Obbarius A, van Maasakkers L, Baer L, et al. Standardization of health outcomes assessment for depression and anxiety: recommendations from the ICHOM Depression and Anxiety Working Group. *Qual Life Res*. 2017;26(12):3211–3225.
4. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998;36(1):8–27.
5. Klabunde CN, Harlan LC, Warren JL. Data sources for measuring comorbidity: a comparison of hospital records and Medicare claims for cancer patients. *Med Care*. 2006;44(10):921–928.
6. Klabunde CN, Warren JL, Legler JM. Assessing comorbidity using claims data: an overview. *Med Care*. 2002;40(suppl 8):IV-26-35.
7. de Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity: a critical review of available methods. *J Clin Epidemiol*. 2003;56(3):221–229.
8. Iezzoni LI, Foley SM, Daley J, Hughes J, Fisher ES, Heeren T. Comorbidities, complications, and coding bias. Does the number of diagnosis codes matter in predicting in-hospital mortality? *JAMA*. 1992;267(16):2197–2203.
9. Susser SR, McCusker J, Belzile E. Comorbidity information in older patients at an emergency visit: self-report vs. administrative data had poor agreement but similar predictive validity. *J Clin Epidemiol*. 2008;61(5):511–515.
10. Preen DB, Holman CDAJ, Lawrence DM, Baynham NJ, Semmens JB. Hospital chart review provided more accurate comorbidity information than data from a general practitioner survey or an administrative database. *J Clin Epidemiol*. 2004;57(12):1295–1304.
11. Katz JN, Chang LC, Sangha O, Fossel AH, Bates DW. Can comorbidity be measured by questionnaire rather than medical record review? *Med Care*. 1996;34(1):73–84.
12. Olomu AB, Corser WD, Stommel M, Xie Y, Holmes-Rovner M. Do self-report and medical record comorbidity data predict longitudinal functional capacity and quality of life health outcomes similarly? *BMC Health Serv Res*. 2012;12(1):398.
13. Katherine F, Amy T, Nicole B, Sigrid C, Andrew JV. Comparison of physician-documented versus patient-reported collection of comorbidities among patients with prostate cancer upon first visit to the urology clinic. *JCO Clin Cancer Inform*. 2018;(2):1–10.
14. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.
15. Bramer WM, Rethlefsen ML, Mast F, Kleijnen J. Evaluation of a new method for librarian-mediated literature searches for systematic reviews. *Res Synth Methods*. 2018;9(4):510–520.
16. Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Syst Rev*. 2017;6(1):245.
17. Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. *J Med Libr Assoc*. 2016;104(3):240–243.
18. Altman DG. Practical Statistics for Medical Research. Boca Raton, FL: CRC Press; 1990.
19. Akoglu H. User's guide to correlation coefficients. *Turk J Emerg Med*. 2018;18(3):91–93.

20. Gliklich RE, Leavy MB, Karl J, Campion DM, Levy D, Berliner E. A framework for creating standardized outcome measures for patient registries. *J Comp Eff Res*. 2014;3(5):473–480.
21. McCusker J, Verdon J, Tousignant P, de Courval LP, Dendukuri N, Belzile E. Rapid emergency department intervention for older people reduces risk of functional decline: results of a multicenter randomized trial. *J Am Geriatr Soc*. 2001;49(10):1272–1281.
22. Corser W, Sikorskii A, Olomu A, Stommel M, Proden C, Holmes-Rovner M. Concordance between comorbidity data from patient self-report interviews and medical record documentation. *BMC Health Serv Res*. 2008;8:85.
23. Holmes-Rovner M, Stommel M, Corser WD, et al. Does outpatient telephone coaching add to hospital quality improvement following hospitalization for acute coronary syndrome? *J Gen Intern Med*. 2008;23(9):1464–1470.
24. Ng X, Low AH, Thumboo J. Comparison of the Charlson Comorbidity Index derived from self-report and medical record review in Asian patients with rheumatic diseases. *Rheumatol Int*. 2015;35(12):2005–2011.
25. Habbous S, Chu KP, Harland LT, et al. Validation of a one-page patient-reported Charlson comorbidity index questionnaire for upper aerodigestive tract cancer patients. *Oral Oncol*. 2013;49(5):407–412.
26. Chaudhry S, Jin L, Meltzer D. Use of a self-report-generated Charlson Comorbidity Index for predicting mortality. *Med Care*. 2005;43(6):607–615.
27. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The self-administered comorbidity questionnaire: a new method to assess comorbidity for clinical and health services research. *Arthritis Rheum.* 2003;49(2):156–163.
28. Stolwijk C, van Tubergen A, Ramiro S, et al. Aspects of validity of the self-administered comorbidity questionnaire in patients with ankylosing spondylitis. *Rheumatology* (Oxford). 2014;53(6):1054–1064.
29. Spoorenberg A, van der Heijde D, de Klerk E, et al. Relative value of erythrocyte sedimentation rate and C-reactive protein in assessment of disease activity in ankylosing spondylitis. *J Rheumatol*. 1999;26(4):980–984.
30. Robinski M, Strich F, Mau W, Girndt M. Validating a patient-reported comorbidity measure with respect to quality of life in end-stage renal disease. *PLoS One*. 2016;11(6):e0157506.
31. Robinski M, Mau W, Lamprecht J, Krauth C, Girndt M. The Choice of Renal Replacement Therapy (CORETH) project: study design and methods. *Clin Kidney J*. 2014;7(6):575–581.
32. Bayliss EA, Ellis JL, Steiner JF. Subjective assessments of comorbidity correlate with quality of life health outcomes: initial validation of a comorbidity assessment instrument. *Health Qual Life Outcomes*. 2005;3:51.
33. Poitras ME, Fortin M, Hudon C, Haggerty J, Almirall J. Validation of the disease burden morbidity assessment by self-report in a French-speaking population. *BMC Health Serv Res*. 2012;12:35.
34. Wijers IGM, Ayala A, Rodriguez-Blazquez C, Rodriguez-Laso A, RodriguezRodriguez V, Forjaz MJ. Disease burden morbidity assessment by self-report: psychometric properties in older adults in Spain. *Geriatr Gerontol Int.* 2017;17(7):1102–1108.
35. Teófilo Rodríguez J, González Cabezas AN, Díaz Veiga P, Rodríguez Rodríguez V. Estudio Longitudinal Envejecer en España: El proyecto ELES. *Perfiles y Tendencias*. 2011;50:1–44.
36. Simpson CF, Boyd CM, Carlson MC, Griswold ME, Guralnik JM, Fried LP. Agreement between self-report of disease diagnoses and medical record validation in disabled older women: factors that modify agreement. *J Am Geriatr Soc*. 2004;52(1):123–127.

37. Kasper JD, Shapiro S, Guralnik JM, Bandeen-Roche KJ, Fried LP. Designing a community study of moderately to severely disabled older women: the Women's Health and Aging Study. *Ann Epidemiol*. 1999;9(8):498–507.
38. Crabtree HL, Gray CS, Hildreth AJ, O'Connell JE, Brown J. The comorbidity symptom scale: a combined disease inventory and assessment of symptom severity. *J Am Geriatr Soc*. 2000;48(12):1674–1678.
39. De-loyde KJ, Harrison JD, Durcinoska I, Shepherd HL, Solomon MJ, Young JM. Which information source is best? Concordance between patient report, clinician report and medical records of patient co-morbidity and adjuvant therapy health information. *J Eval Clin Pract*. 2015;21(2):339–346.
40. Young JM, Butow PN, Walsh J, et al. Multicenter randomized trial of centralized nurse-led telephone-based care coordination to improve outcomes after surgical resection for colorectal cancer: the CONNECT intervention. *J Clin Oncol.* 2013;31(28):3585–3591.
41. Gad BV, Higuera CA, Klika AK, Elsharkawy KA, Barsoum WK. Validity of patient-reported comorbidities before total knee and hip arthroplasty in patients older than 65 years. *J Arthroplasty.* 2012;27(10):1750–1756.e1s.
42. Hansen H, Schafer I, Schon G, et al. Agreement between self-reported and general practitioner-reported chronic conditions among multimorbid patients in primary care - results of the MultiCare Cohort Study. *BMC Fam Pract*. 2014;15:39.
43. Schafer I, Hansen H, Schon G, et al. The German MultiCare-study: patterns of multimorbidity in primary health care - protocol of a prospective cohort study. *BMC Health Serv Res*. 2009;9:145.
44. Horton M, Rudick RA, Hara-Cleaver C, Marrie RA. Validation of a self-report comorbidity questionnaire for multiple sclerosis. *Neuroepidemiology*. 2010;35(2):83–90.
45. Iecovich E, Biderman A. Concordance between self-reported and physicianreported chronic co-morbidity among disabled older adults. *Can J Aging*. 2013;32(3):287–297.
46. Potosky AL, Harlan LC, Stanford JL, et al. Prostate cancer practice patterns and quality of life: the Prostate Cancer Outcomes Study. *J Natl Cancer Inst.* 1999;91(20):1719–1724.
47. Boissonnault WG, Badke MB. Collecting health history information: the accuracy of a patient self-administered questionnaire in an orthopedic outpatient setting. *Phys Ther*. 2005;85(6):531–543.
48. Fan VS, Au D, Heagerty P, Deyo RA, McDonell MB, Fihn SD. Validation of casemix measures derived from self-reports of diagnoses and health. *J Clin Epidemiol*. 2002;55(4):371–380.
49. Fihn SD, McDonell MB, Diehr P, et al. Effects of sustained audit/feedback on self-reported health status of primary care patients. *Am J Med*. 2004;116(4):241–248.
50. Lorem GF, Schirmer H, Emaus N. Health Impact Index. Development and validation of a method for classifying comorbid disease measured against self-reported health. *PLoS One*. 2016;11(2):e0148830.
51. Jacobsen BK, Eggen AE, Mathiesen EB, Wilsgaard T, Njolstad I. Cohort profile: the Tromso Study. *Int J Epidemiol*. 2012;41(4):961–967.
52. Lucke T, Herrera R, Wacker M, et al. Systematic analysis of self-reported comorbidities in large cohort studies - a novel stepwise approach by evaluation of medication. *PLoS One*. 2016;11(10):e0163408.
53. Karch A, Vogelmeier C, Welte T, et al. The German COPD cohort COSYCONET: aims, methods and descriptive analysis of the study population at baseline. *Respir Med*. 2016;114:27–37.

54. Md Yusof MY, Horan MA, Jones M, McInnes L, Rabbitt PM, Pendleton N. Developing a self-reported comorbidity index to predict mortality of community-dwelling older adults. *Arch Gerontol Geriatr*. 2010;50(3):e63–67.
55. Merkin SS, Cavanaugh K, Longenecker JC, Fink NE, Levey AS, Powe NR. Agreement of self-reported comorbid conditions with medical and physician reports varied by disease among end-stage renal disease patients. *J Clin Epidemiol*. 2007;60(6):634–642.
56. Powe NR, Klag MJ, Sadler JH, et al. Choices for healthy outcomes in caring for end stage renal disease. *Seminars in Dialysis*. 1996;9(1):9–11.
57. Mukerji SS, Duffy SA, Fowler KE, Khan M, Ronis DL, Terrell JE. Comorbidities in head and neck cancer: agreement between self-report and chart review. *Otolaryngol Head Neck Surg*. 2007;136(4):536–542.
58. Paleri V, Wight RG. A cross-comparison of retrospective notes extraction and combined notes extraction and patient interview in the completion of a comorbidity index (ACE-27) in a cohort of United Kingdom patients with head and neck cancer. *J Laryngol Otol*. 2002;116(11):937–941.
59. Pati S, Hussain MA, Swain S, et al. Development and validation of a questionnaire to assess multimorbidity in primary care: an Indian experience. *Biomed Res Int*. 2016;2016:6582487.
60. Lee SJ, Lindquist K, Segal MR, Covinsky KE. Development and validation of a prognostic index for 4-year mortality in older adults. *JAMA*. 2006;295(7):801–808.
61. The Health and Retirement Survey. Ann Arbor: University of Michigan; 1996. http://hrsonline.isr.umich.edu/docs/sample/sho_samp.php?hfyle=ref023b&xtyp=1. Accessed January 19, 2006.
62. The Health and Retirement Survey. Ann Arbor: University of Michigan. http://hrsonline.isr.umich.edu/intro/sho_uinfo.php?hfyle=overview&xtyp=2. Accessed January 19, 2006.
63. Voaklander DC, Kelly KD, Jones CA, Suarez-Almazor ME. Self-report comorbidity and health related quality of life – a comparison with record based co-morbidity measures. *Social Indicators Research*. 2004;66(3):213–228.
64. Selim AJ, Fincke G, Ren XS, et al. Comorbidity assessments based on patient report: results from the Veterans Health Study. *J Ambul Care Manage*. 2004;27(3):281–295.
65. Vigen C, Kwan ML, John EM, et al. Validation of self-reported comorbidity status of breast cancer patients with medical records: the California Breast Cancer Survivorship Consortium (CBCSC). *Cancer Causes Control.* 2016;27(3):391–401.
66. John EM, Phipps AI, Davis A, Koo J. Migration history, acculturation, and breast cancer risk in Hispanic women. *Cancer Epidemiol Biomarkers Prev*. 2005;14(12):2905–2913.
67. Caan B, Sternfeld B, Gunderson E, Coates A, Quesenberry C, Slattery ML. Life After Cancer Epidemiology (LACE) Study: a cohort of early stage breast cancer survivors (United States). *Cancer Causes Control.* 2005;16(5):545–556.
68. Huntley AL, Johnson R, Purdy S, Valderas JM, Salisbury C. Measures of multimorbidity and morbidity burden for use in primary care and community settings: a systematic review and guide. *Ann Fam Med*. 2012;10(2):134–141.
69. Violán C, Foguet-Boreu Q, Hermosilla-Pérez E, et al. Comparison of the information provided by electronic health records data and a population health survey to estimate prevalence of selected health conditions and multimorbidity. *BMC Public Health*. 2013;13(1):251.
70. Sprangers MA, Cull A, Bjordal K, Groenvold M, Aaronson NK. The European Organization for Research and Treatment of Cancer. Approach to quality of life assessment: guidelines

for developing questionnaire modules. EORTC Study Group on Quality of Life. *Qual Life Res*. 1993;2(4):287–295.

71. Farnik M, Pierzchała WA. Instrument development and evaluation for patient-related outcomes assessments. *Patient Relat Outcome Meas*. 2012;3:1–7.
72. Leroy L, Bayliss E, Domino M, et al. The Agency for Healthcare Research and Quality Multiple Chronic Conditions Research Network: overview of research contributions and future priorities. *Med Care*. 2014;52(Suppl 3):S15–S22.
73. Gliklich RE DN, Leavy MB, eds. Registries for evaluating patient outcomes: a user's guide [Internet]. 3rd ed. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014 Apr. 1. *Patient Registries*. https://www.ncbi.nlm.nih.gov/books/NBK208643/. Accessed July 24, 2019.

SUPPLEMENTAL MATERIAL

Supplementary Search Strategy: Literature Search Terms

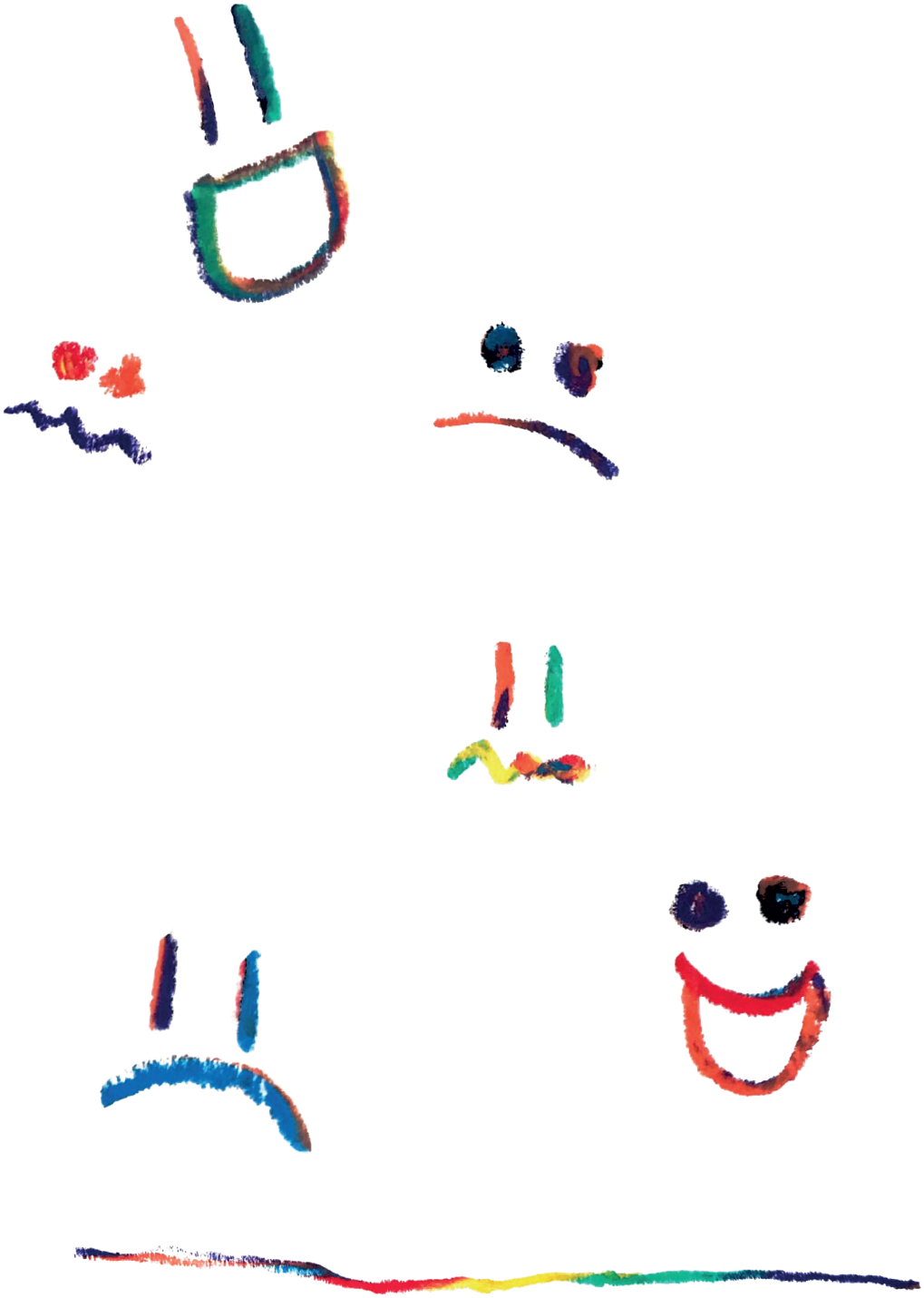
Embase.com – ('comorbidity'/mj OR 'comorbidity assessment'/mj/exp OR 'multiple chronic conditions'/mj OR (comorbid* OR co-morbid* OR multimorbid* OR (multi* NEAR/3 (morbid* OR chronic*))):ab,ti) AND ('surveys and questionnaires'/mj/exp OR 'self report'/mj OR 'patient-reported outcome'/mj OR (((patient* OR self) NEXT/1 (report* OR based OR generate* OR rate OR rating OR rated)) OR ((patient* OR self) NEAR/3 (questionnaire* OR survey* OR assess* OR index OR indices OR instrument*))):ti) AND ('validation process'/exp OR 'validation study'/de OR 'validity'/exp OR 'reliability'/exp OR 'reproducibility'/exp OR (validat* OR validit* OR ((develop* OR value* OR limitation* OR feasib* OR valid OR reliab*) NEAR/10 (questionnaire* OR instrument* OR version* OR tool*))):ab,ti) NOT ([Conference Abstract]/lim OR [Letter]/lim OR [Note]/lim OR [Editorial]/lim) AND [english]/lim

Medline Ovid – (exp * Comorbidity/ OR (comorbid* OR co-morbid* OR multimorbid* OR (multi* ADJ3 (morbid* OR chronic*))).ab,ti.) AND (* Surveys and Questionnaires/ OR * Self Report/ OR * Patient Reported Outcome Measures/ OR (((patient* OR self) ADJ (report* OR based OR generate* OR rate OR rating OR rated)) OR ((patient* OR self) ADJ3 (questionnaire* OR survey* OR assess* OR index OR indices OR instrument*))).ti.) AND (Validation Studies as Topic/ OR Validation Studies/ OR Reproducibility of Results/ OR (validat* OR validit* OR ((develop* OR value* OR limitation* OR feasib* OR valid OR reliab*) ADJ10 (questionnaire* OR instrument* OR version* OR tool*))).ab,ti.) NOT (letter* OR news OR comment* OR editorial* OR congres* OR abstract* OR book* OR chapter* OR dissertation abstract*).pt. AND english.la.

Cochrane CENTRAL – ((comorbid* OR co-morbid* OR multimorbid* OR (multi* NEAR/3 (morbid* OR chronic*))):ab,ti) AND ((((patient* OR self) NEXT/1 (report* OR based OR generate* OR rate OR rating OR rated)) OR ((patient* OR self) NEAR/3 (questionnaire* OR survey* OR assess* OR index OR indices OR instrument*))):ti) AND ((validat* OR validit* OR ((develop* OR value* OR limitation* OR feasib* OR valid OR reliab*) NEAR/10 (questionnaire* OR instrument* OR version* OR tool*))):ab,ti)

Google Scholar – comorbidity|"co|multi morbidity"|multimorbid "patient|self reported|based|generated|rated|questionnaire|survey|assessment|index|indices|instrument" validation|validity|development|reliability

Web of Science – ((TS=(comorbid* OR co-morbid* OR multimorbid* OR (multi* NEAR/2 (morbid* OR chronic*)))) AND (TI=(((patient* OR self) NEAR/1 (report* OR based OR generate* OR rate OR rating OR rated)) OR ((patient* OR self) NEAR/2 (questionnaire* OR survey* OR assess* OR index OR indices OR instrument*))))) AND TS=((validat* OR validit* OR ((develop* OR value* OR limitation* OR feasib* OR valid OR reliab*) NEAR/10 (questionnaire* OR instrument* OR version* OR tool*)))) AND DT=(article) AND LA=(english)



Chapter 3 Value-Based Healthcare in Ischemic Stroke Care: Case-Mix Adjustment Models for Clinical and Patient-Reported Outcomes

Arvind Oemrawsingh, MD MHS[1], Nikki van Leeuwen, PhD[1], Esmee Venema, MD MHS[1,7], Martien Limburg, MD PhD[2,10], Frank-Erik de Leeuw MD, PhD[3], Markus P. Wijffels, MD[4], Aafke J. de Groot, MD[5], Pieter H.E. Hilkens, MD PhD[6], Jan A. Hazelzet, MD PhD[1], Diederik W.J. Dippel, MD PhD[7], Carla H. Bakker, PhD[8], Helene R. Voogdt-Pruis, PhD[9,10], Hester F. Lingsma, PhD[1]

[1] *Center for Medical Decision Making, Department of Public Health, Erasmus University Medical Center Rotterdam, PO Box 2040, 3000 CA, Rotterdam, the Netherlands*
[2] *Department of Neurology, Flevoziekenhuis, Almere, the Netherlands*
[3] *Department of Neurology, Donders Institute for Brain Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands*
[4] *Department of Neurorehabilitation, Rijndam Rehabilitation, Rotterdam, the Netherlands*
[5] *Department of General Practice and Elderly Care Medicine, Amsterdam University Medical Center, Amsterdam, the Netherlands / Vivium Naarderheem, Naarden, the Netherlands*
[6] *Department of Neurology, St. Antonius Hospital, Nieuwegein, the Netherlands*
[7] *Department of Neurology, Erasmus University Medical Center, Rotterdam, the Netherlands*
[8] *Expert Centre Quality Registries, Leiden University Medical Center, Leiden, the Netherlands*
[9] *EnCorps, Hilversum, the Netherlands*
[10] *Stroke Knowledge Network Netherlands, Utrecht, the Netherlands*

BMC Medical Research Methodology, 2019 Dec 5;19(1):229.

ABSTRACT

Background

Patient-Reported Outcome Measures (PROMs) have been proposed for benchmarking health care quality across hospitals, which requires extensive case-mix adjustment. The current study's aim was to develop and compare case-mix models for mortality, a functional outcome, and a patient-reported outcome measure (PROM) in ischemic stroke care.

Methods

Data from ischemic stroke patients, admitted to four stroke centers in the Netherlands between 2014 and 2016 with available outcome information (N = 1022), was analyzed. Case-mix adjustment models were developed for mortality, modified Rankin Scale (mRS) scores and EQ-5D index scores with respectively binary logistic, proportional odds and linear regression models with stepwise backward selection. Predictive ability of these models was determined with R-squared (R2) and area-under-the-receiver-operating-characteristic-curve (AUC) statistics.

Results

Age, NIHSS score on admission, and heart failure were the only common predictors across all three case-mix adjustment models. Specific predictors for the EQ-5D index score were sex ($\beta = 0.041$), socio-economic status ($\beta = -0.019$) and nationality ($\beta = -0.074$). R2-values for the regression models for mortality (5 predictors), mRS score (9 predictors) and EQ-5D utility score (12 predictors), were respectively R2 = 0.44, R2 = 0.42 and R2 = 0.37.

Conclusions

The set of case-mix adjustment variables for the EQ-5D at three months differed considerably from the set for clinical outcomes in stroke care. The case-mix adjustment variables that were specific to this PROM were sex, socio-economic status and nationality. These variables should be considered in future attempts to risk-adjust for PROMs during benchmarking of hospitals.

Keywords

Ischemic Stroke; Case-Mix; Risk Adjustment Model; Patient-Reported Outcome Measure; Value-Based Healthcare

BACKGROUND

The growing trend to benchmark certain health care performance indicators – to assess the health care quality between institutions – requires careful consideration of the methodology that is being used.[1, 2] Healthcare is evolving towards a value-based healthcare framework with more emphasis on Patient-Reported Outcome Measures (PROMs), that will not only facilitate opportunities for performance improvement at an individual patient level when these measures are used in clinical practice, but may also be useful for benchmarking across providers.[3, 4, 5] PROMs can be defined as feedback directly from the individual patient on their own health condition (e.g. symptoms and health-related quality of life), thus without external interpretation.[6] PROMs can be either disease-specific (e.g. Neuro-QOL[7]) or generic (e.g. EQ-5D[8]).

An important consideration for meaningful comparisons across hospitals is the case-mix adjustment of the patient populations for each health care provider.[9, 10] By adjusting for the heterogeneity of patient characteristics in inter-hospital comparisons, a larger part of the estimated variation between hospital performances will be attributable to the quality of care provided to patients rather than factors outside of the healthcare providers' control.

In stroke, the most commonly used clinical outcome measures are mortality and the modified Rankin Scale (mRS). There has been considerable research conducted on prognostic models for these clinical outcomes, which also encompass variables for case-mix adjustment.[11] Although there is a strong trend to use PROMs for benchmarking purposes[12], there still remains a lack of case-mix models to predict patient-reported outcomes as compared to clinical outcomes.[11] The aim of this study was to identify the specific variables for case-mix adjustment for a generic PROM (EQ-5D) and compare them to case-mix variables for clinical outcomes in acute ischemic stroke.

METHODS

Patient Population and Data Collection

A core set of baseline patient characteristics, performance indicators and outcome measures were registered from March 2014 till August 2016 of four stroke care centers in the Netherlands, of which 1 was a university and 3 were district-based hospitals. The original database contained consecutive acutely admitted ischemic

stroke patients of which demographic, process indicators and outcome measures were registered.

The three outcome measures were mortality at 3 months, modified Rankin Scale (mRS) score at 3 months and EuroQol-5D index score at 3 months. The mRS is a commonly used clinician-reported scale, which measures the degree of disability after a stroke, with scores ranging from 0 to 6 (Figure 1).[13] The mRS score at 3 months post-discharge was generally recorded by trained nurses, either by phone or at the outpatient clinic. The EQ-6D, a generic health-related quality of life (HRQOL) instrument, is based on the EQ-5D (dimensions: usual activities, self-care or autonomy, mobility, pain/ discomfort, and anxiety/ depression) with an additional question on cognitive functioning. The survey has been translated into Dutch and validated in previously published literature.[8, 14, 15] The post-discharge EQ-6D data was captured through either face-to-face or telephone interviews with patients themselves or their proxies. Due to the lack of a validated EQ-6D index tariff, the utility score was derived and transformed through the EQ-5D index tariff, by ignoring the "cognitive" dimension of the EQ-6D.[16,17,18] This EQ-5D tariff is an algorithm for attaching values to all 3125 health states often used in economic evaluations. This utility score can be used to compare to population norms or to calculate quality-adjusted life years (QALY's).[16, 17, 19] The authors will, for the remainder of this article, solely mention "EQ-5D index score" as the patient-reported outcome of interest to avoid any confusion. This EQ-5D index score ranged from 0 (death) to 1 (perfect health) and signified the patient's perspective on his/ her own health. Because there still is no consensus on the minimal clinically important difference on the EQ-5D utility score in stroke populations[20], it was decided to keep the EQ-5D utility score as a continuous outcome rather than modify it to an ordinal outcome based on arbitrary cut-offs.

Score	Definition
0	No symptoms
1	No significant disability despite some symptoms; able to carry out all usual duties and activities
2	Slight disability; unable to carry out all previous activities, but able to look after own affairs without assistance
3	Moderate disability; requiring some help, but able to walk without assistance
4	Moderate severe disability; unable to walk without assistance and unable to attend to own bodily needs without assistance
5	Severe disability; bedridden, incontinent and requiring constant nursing care and attention
6	Death

Figure 1 | Modified Rankin Scale (mRS).

Missing baseline patient characteristics (among which case-mix variables) were imputed 10 times using multiple imputation in the original database (N = 2733 patients of 4 stroke centers), assuming missingness at random. Predictors (including stroke center) and outcomes also served as indicators for the imputation model.[21] Figure 2 showcases the substantial proportion of missing mRS and EQ-5D data that were filtered out before the three case-mix adjustment models were developed. Thus, all three regression models were developed using an imputed (10 iterations) dataset also containing the "original" 1022 patients. Potential reasons for missing patient-reported outcome data were patients being too sick to fill questionnaires out, and loss-to-follow at 3 months (patients being unreachable due to staying at a nursing home/ rehabilitation center, or because of their tremendous recovery).

Figure 2 | Flowchart of Study Population Selection.

Case-Mix Models

Patient characteristics that could differ between hospitals and could be predictive of outcomes were considered potential candidate case-mix variables and were identified based on clinical experience and past literature. Those included: age, sex, nationality, socio-economic status (SES), smoking, cardiovascular comorbidity (e.g. hypertension, hyperlipidemia), stroke in past history, diabetes, cancer, connective tissue disease, Charlson comorbidity index (CCI), stroke onset-door time, initial National Institutes of Health Stroke Scale (NIHSS) score and the presence of a caregiver. The SES was generated by the ranking of status scores (based on neighborhoods/ zip codes) that have been calculated and published by Social and Cultural Planning Office (SCP), a Dutch governmental institution.[22, 23]

Statistical Analysis

Descriptive statistics were presented as counts (percentages) or median ± interquartile range (IQR). Nonparametric tests were used where appropriate to deter-

mine (unadjusted) differences in patient populations between the four healthcare providers, using the Pearson chi-squared statistic for categorical variables and the Kruskal-Wallis test for continuous variables. A p-value < 0.05 was considered significant. To assess the adjusted effect of the potential case-mix independent variables, the models were developed using logistic, ordinal and linear regression models respectively for mortality, mRS score and EQ-5D index score with stepwise backward selection. This stepwise (regression-based) method initially tests all the predictors in a regression model and subsequently eliminates the least significant variables in a stepwise approach with a certain cut-off p-value.[21] In this study, the AIC (Akaike information criterion)[24], equivalent to $p < 0.157$, was used as a criterion.

For the logistic and ordinal regression models, the odds ratios (ORs) with 95% confidence intervals (CI) were calculated per predictor. Beta's (ß's) and 95% CI were calculated for predictors in the linear regression model. The ß coefficient indicates the change in outcome (units on the EQ-5D index score scale) for one unit change in the predictor variable. The ability of the case-mix models to explain the variability ('goodness-of-fit') of these 3 outcomes was expressed by calculating the R^2 (R-squared) statistics.[25] The predictors for each model were added in consecutive order based on the p-values (lowest to highest p-value) and coefficients. The explained variance of each additional predictor was demonstrated till each model was completed.

Because the R^2 measure is not immediately comparable between different regression models (logistic vs. ordinal vs. linear), the AUC (area-under-the receiver-operating-characteristic-curve) statistic was also included to get a sense of the comparability between the three risk-adjustment models. For this additional analysis, both mRS (0–2 vs. 3–6) and EQ-5D ($< 0.$ 65 vs. ≥ 0.65) were transformed to a binary outcome variable in order to compare the three logistic regression models. The EQ-5D index score of 0.65 was chosen as a cut-off value, as it was the estimated median score in this study sample. The statistical analysis was carried out by using IBM SPSS Statistics 21 & RStudio version 1.0.153.0-© 2009–2016 RStudio, Inc. software.

RESULTS

Patient Characteristics

In total, 1022 ischemic stroke patients were included. The number of patients per studied hospital varied from 29 to 555 patients. 57% of total study participants were men (Table 1). The unadjusted median age was significantly different (range

Table 1 | Characteristics of all Ischemic Stroke Patients (N =1022) and per Dutch Stroke Center, Admitted from March 2014 – August 2016 in Four Dutch Stroke Hospitals.

Patient characteristics	All patients (N = 1022)	Stroke Center I, University (N = 222)	Stroke Center II, District-based (N = 555)	Stroke Center III, District-based (N = 216)	Stroke Center IV, District-based (N = 29)	P-value	Missing data, N (%)
Male, N (%)	578 (57)	139 (63)	315 (57)	109 (51)	15 (52)	0.076	
Age, median (IQR)	74 (64 – 82)	70 (59 – 80)	76 (66 – 83)	72 (63 – 82)	78 (72 – 85)	0.001	
Nationality						0.351	84 (8)
Native Dutch	888 (95)	171 (95)	517 (94)	174 (97)	26 (93)		
Foreigner	50 (5)	9 (5)	34 (6)	5 (3)	2 (7)		
Smoking, N (%)	225 (22)	51 (24)	131 (25)	37 (20)	6 (23)	0.604	72 (7)
SES, N (%)						<0.001	8 (1)
Low	335 (33)	101 (46)	131 (24)	81 (38)	22 (82)		
Middle	427 (42)	94 (43)	214 (39)	119 (55)	0 (0)		
High	252 (25)	25 (11)	207 (38)	15 (7)	5 (19)		
NIHSS on admission, median (IQR)	4 (2 – 12)	5 (2 – 9)	4 (2 – 14)	3 (1 – 8)	3 (2 – 14)	0.028	75 (7)
CCI, median (IQR)	1 (0 – 2)	1 (0 – 2)	1 (0 – 2)	1 (0 – 2)	1 (0 – 3)	0.133	89 (9)
Comorbidities, N (%)							
Hypertension	546 (53)	123 (56)	332 (60)	74 (35)	17 (59)	<0.001	9 (1)
Myocardial infarction	103 (10)	15 (7)	71 (13)	17 (8)	0 (0)	0.012	15 (2)
Heart failure	66 (7)	2 (1)	52 (9)	8 (4)	4 (15)	<0.001	22 (2)
Previous stroke/ TIA	274 (27)	58 (26)	156 (28)	54 (26)	6 (21)	0.745	10 (1)
Carotid stenosis	64 (6)	11 (5)	44 (8)	8 (4)	1 (4)	0.093	50 (5)
PAOD	80 (8)	13 (6)	46 (8)	17 (8)	4 (15)	0.374	14 (1)
Diabetes mellitus	39 (4)	9 (4)	25 (5)	4 (2)	1 (3)	0.409	15 (2)
Connective tissue disease	26 (3)	0 (0)	13 (2)	13 (7)	0 (0)	<0.001	24 (2)
Cancer	112 (11)	32 (14)	44 (8)	34 (16)	2 (7)	0.004	6 (1)
Metastasis	24 (2)	8 (4)	10 (2)	5 (3)	1 (4)	0.514	27 (3)
Caregiver post-discharge, N (%)	563 (55)	140 (69)	292 (61)	121 (68)	10 (83)	0.066	150 (15)
Onset-to-door time, median minutes (IQR)	213 (80- 672)	275 (72 – 959)	215 (93 – 641)	149 (63 – 384)	113 (41 – 368)	0.002	153 (15)

Note. IQR, inter-quartile range; *SES*, socio-economic status, derived from status scores based on zip codes; *NIHSS* National Institute of Health Stroke Scale; *CCI*, Charlson Comorbidity Index; *TIA*, transient ischemic attack; *PAOD*, peripheral arterial occlusive disease.

70–78 years, p = 0.001) across the four stroke care centers. Most patients (87%) were native Dutch inhabitants. Both the Charlson Comorbidity Index and the stroke rate in patient history were similar across the 4 stroke patient cohorts. There was a significant unadjusted difference (range 113–275 minutes, p = 0.002) in the onset-to-door time between the patient populations of the four stroke care centers.

The 3-month mortality was 24.5% (Table 2). The unadjusted difference in 3-month mortality rates between the four stroke centers was significant (p < 0.001). 581 (57%) of all patients had a favorable degree of disability (mRS < 3). There was also a significant unadjusted difference in mRS scale scores between the four stroke centers. The median EQ-5D index score at 3 months for all patients was 0.65 (inter-quartile range 0.10–0.83), and the unadjusted difference across the four stroke centers was also significant (p < 0.001). Missing mRS outcomes were 1205/2733 (44.1%) in the original database, with most missing mRS data being observed in stroke center IV (192/238 = 80.7%) (data not shown).

Table 2 | Outcome Measures of Ischemic Stroke Patients (N = 1022).

Outcome variables	All patients (N = 1022)	Stroke Center I, University (N = 222)	Stroke Center II, District-based (N = 555)	Stroke Center III, District-based (N = 216)	Stroke Center IV, District-based (N = 29)	P-value
mRS at 3 months, N (%)						<0.001
0	89 (8.7)	37 (16.7)	25 (4.5)	27 (12.5)	0	
1	262 (25.6)	62 (27.9)	143 (25.8)	53 (24.5)	4 (13.8)	
2	230 (22.5)	37 (16.7)	130 (23.4)	58 (26.9)	5 (17.2)	
3	108 (10.6)	23 (10.4)	67 (12.1)	15 (6.9)	3 (10.3)	
4	68 (6.7)	24 (10.8)	32 (5.8)	11 (5.1)	1 (3.4)	
5	15 (1.5)	2 (0.9)	11 (2.0)	2 (0.9)	0	
6	250 (24.5)	37 (16.7)	147 (26.5)	50 (23.1)	16 (55.2)	
EQ-5D index score at 3 months, median (IQR)	0.65 (0.1 – 0.83)	0.781 (0.45 – 1.00)	0.61 (0 – 0.78)	0.65 (0.28 – 0.83)	0 (0 – 0.60)	<0.001

Note. mRS, Modified Rankin Scale; *EQ-5D*, EuroQol 5-Dimension; *IQR,* inter-quartile range

Case-Mix Models

Table 3 shows the remaining predictors in the regression models after backward selection for mortality, mRS and EQ-5D utility scores. The "strongest" (based on lowest p-values) independent variables in the model for mortality were age (OR = 1.07), NIHSS score on admission (OR = 1.17) and the Charlson's comorbidity index (OR = 1.22). The strongest predictors for mRS at 3 months were age (OR = 1.04),

Table 3 | Case-Mix Risk Adjustment Models for Mortality, mRS and EQ-5D.

	Model 1*: Mortality at 3 months		Model 2*: mRS score at 3 months		Model 3*: EQ-5D index score at 3 months	
(Nagelkerke) R^2	0.44		0.42		0.37	
AUC°	0.87		0.83		0.78	
Variables	**Multivariable OR (95% CI)**	**P-value**	**Multivariable OR (95% CI)**	**P-value**	**ßeta (SE)**	**P-value**
Age	1.07 (1.05 – 1.09)	<0.001	1.04 (1.03 – 1.05)	<0.001	-0.007 (0.001)	<0.001
NIHSS on admission	1.17 (1.14 – 1.20)	<0.001	1.17 (1.14 – 1.19)	<0.001	-0.020 (0.001)	<0.001
Heart failure	2.54 (1.30 – 4.97)	0.007	3.58 (2.14 – 5.98)	<0.001	-0.130 (0.042)	0.002
Previous stroke			1.74 (1.33 – 2.27)	<0.001	-0.063 (0.023)	0.007
Smoking			1.58 (1.18 – 2.13)	0.003	-0.055 (0.025)	0.028
Caregiver at discharge			0.67 (0.51 – 0.86)	0.002	0.057 (0.022)	0.010
Connective tissue disease			2.07 (0.93 – 4.61)	0.074	-0.119 (0.061)	0.053
Cancer			1.99 (1.35 – 2.94)	0.001	-0.065 (0.042)	0.122
CCI	1.22 (1.13 – 1.32)	<0.001			-0.020 (0.007)	0.003
Onset-to-door time (per 10 minutes)	1.00 (1.00 – 1.00)	0.123				
Diabetes			2.09 (1.12 – 3.88)	0.020		
Sex (Ref: Female)					0.041 (0.020)	0.036
SES					-0.019 (0.010)	0.053
Nationality (Ref: Native Dutch)					-0.074 (0.045)	0.097

Note. SES, socio-economic status, derived from neighborhood-based ranking; *NIHSS*, National Institute of Health Stroke Scale; *CCI*, Charlson Comorbidity Index; R^2 R-squared; *95% CI*, 95% confidence interval; *SE*, standard error
* = The case-mix models was using backward selection with the Akaike Information Criterion (AIC) as a cut-off for the p-value. The imputed dataset (10 imputations of original data N = 1022) was used for the development of all three case-mix model development
[a] = AUC values were calculated for mortality, dichotomized mRS scores (0–2 vs. 3–6) and dichotomized EQ-5D index scores (< 0.65 vs. ≥ 0.65)

NIHSS score at admission (OR = 1.17), heart failure (OR = 3.58) and previous stroke (OR = 1.74). There were only three overlapping predictors for the three different outcomes: age, NIHSS score on admission and heart failure. Exclusive predictors for the EQ-5D index score were sex (β = 0.041), socio-economic status (β = − 0.019), and nationality (β = − 0.074).

The binary logistic regression model for mortality had an $R^2 = 0.44$ (Table 3), compared to the ordinal regression model for the mRS which had an $R^2 = 0.42$, and the linear regression model for the EQ-5D utility score with a $R^2 = 0.37$. The largest increase in R^2 was after the addition of NIHSS to the models for mortality and mRS, and age to the EQ-5D index score model (Figure 3). After mRS and EQ-5D index scores were both transformed to a dichotomous outcome, AUC's were compared

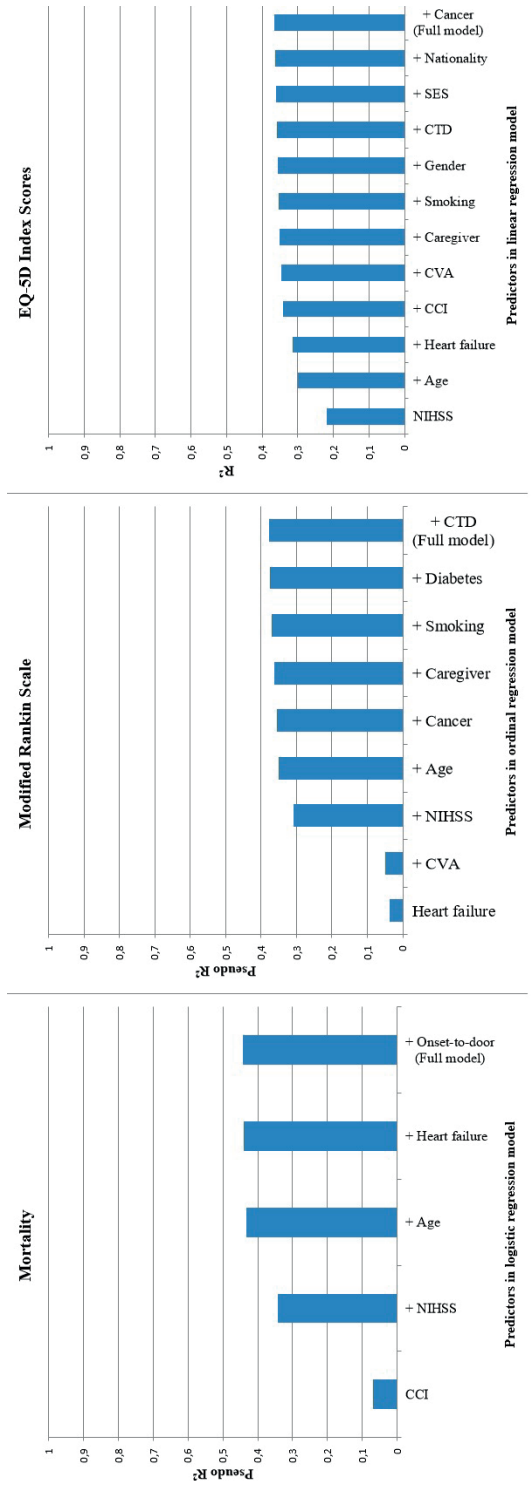


Figure 3 | Prognostic Value of Univariable and Full Models for Three Outcomes, Expressed in Percentage Explained Variance (R2)

between all three binary logistic regression models: AUC = 0.87 (mortality) vs. AUC = 0.83 (mRS ≥ 3) vs. AUC = 0.78 (EQ-5D index score ≥ 0. 65). As opposed to the models for mortality and mRS, it took more predictors in the model for EQ-5D index scores for the predictive ability to reach a plateau.

DISCUSSION

The objective of this study was to construct and compare case-mix adjustment models for three different outcomes, of which two were clinical (mortality and modified Rankin Scale at 3 months) and one patient-reported (EQ-5D utility score at 3 months). The three case-mix models had several predictors in common: age, NIHSS score at hospital admission, and heart failure. However, the most important difference in the case-mix adjustment models was that sex, nationality, and socio-economic status remained significant case-mix variables specifically for the PROM in contrast to the models for the clinical outcomes. It has to be stated that even if a predictor is significantly associated with the outcome, it doesn't necessarily have to be included as a case-mix variable, if the prevalence distribution of the variable and its effect on the outcome of interest is similar across hospitals. The R-squared (R^2) statistics of the model for the patient-reported outcome measure (PROM) was somewhat lower in comparison to the R^2 statistics for mortality and the modified Rankin Scale (mRS), but contained more variables.

There have been multiple models previously developed and validated to predict clinical outcomes after stroke.[11] Bray et al.[26] developed and externally validated two case-mix models with 30-day post-stroke mortality as an outcome. The predictors included in the final models were similar to the findings of the current study: age, NIHSS on admission and atrial fibrillation. On the other hand, there has not been much research conducted on the development of case-mix factors for patient-reported outcomes (e.g. EQ-5D) in stroke care.[27] There was some overlap in the remaining case-mix variables in this study and those identified in previously published articles.[28] A review by Carod-Artal et al.[29] identified age, sex, stroke severity, physical impairment, functional status, and mental impairment as predictors for the health-related quality of life (HRQOL) after stroke. Mar et al.[30] also found the male gender and the NIHSS to be significantly associated with better EQ-5D values. The negative association between (history of) cancer and a lower quality of life (lower EQ-5D scores) in this study confirms previously published literature.[23,31,32]

A striking observation is the caregiver presence post-discharge as a statistically significant predictor variable for mRS score at 3 months with an OR = 0.67 (95% CI 0.51–0.86; p = 0.002) and for the EQ-5D utility score at 3 months with a ßeta = 0.057 (SE 0.022). This observation of caregiver presence at hospital discharge being associated with a lower mRS score (better clinical outcome) and a higher EQ-5D utility score (better quality of life) at 3 months, highlights the potential benefits that a caregiver could offer (e.g. patient motivation, facilitating rehabilitative care) leading to improved functional status and quality of life. However, definite conclusions cannot be drawn about this association, because the definition of "caregiver" and "caregiver presence" was not similar across the multiple stroke centers; it was unclear if the absence of a caregiver implied no indication (e.g. low mRS score) or no need (admittance to a revalidation center or nursing home).

Strengths and Limitations

A considerable strength of this study is that it explores a relatively new field of research namely case-mix adjustments for PROMs in order to make inter-hospital performance comparisons. The case-mix variables for a PROM do not imply additional registration burden for recording data in quality registries because general (relevant) demographic variables (e.g. age, gender and socio-economic status) are already captured in standardized fashion. This is a major strength of the study.

An important limitation of this study is the notably large amount of missing outcome (mRS and EQ-5D) data in the original database. This problem is not uncommon in registries that are routinely acquired for the purpose of quality of care assessment, and it was the main reason this study solely focused on the development of case-mix risk adjustment models rather than benchmarking the included stroke centers. Although the estimated regression coefficients of all three case-mix models might be somewhat biased due to the substantial missingness, it is less important in this context and more about the differences in case-mix variables between the models. The missing data issue was partially countered by the use of multiple imputation for the predictor variables. The distribution of patient characteristics was compared between missing and non-missing mRS and EQ-5D groups (data not shown). These analyses showed significant differences in distribution in NIHSS score, SES rank (low, middle, high), nationality, and some cardiovascular comorbidities (hypertension, heart failure, hyperlipidemia) between missing and non-missing mRS and EQ-5D data. This observation implies that the generalizability of the final set of case-mix variables, observed in this study, should be corroborated in future research.

Another limitation is the loss-to-follow-up bias in this stroke registry: if missing 3-month mRS and EQ-5D data could be attributed to either patients' full recovery or an extended stay in a rehabilitation center/ nursery home, it is quite possible that known outcome data could be skewed (both directionalities possible), seeing it was mostly recorded at outpatient clinics. Other stroke registries (e.g. European Safe Implementation of Thrombolysis in Stroke-Monitoring Study (SITS-MOST)[33] and UK Sentinel Stroke National Audit Program (SSNAP)[34]) have also incorporated patient-reported outcomes, which are typically collected at 3–6 months post-discharge through diverse methods like face-to-face interviews, telephone interviews or mailed questionnaires.[35] As the collection of PROMs at these time points can be challenging due to varying post-discharge patient trajectories and/ or substantial resource requirements (personnel and costs), future research should focus on efficient methods to optimally capture PROM data as part of value-based stroke care. This is an essential step that should to be taken before (case-mix) risk adjustment models are further developed.

In this study, R^2 values were compared to pseudo-R^2 values although they are not directly comparable. However, the objective of this study was to showcase the differences in predictors between the three models. It has to be noted that some potentially relevant psychological case-mix variables (e.g. depression, anxiety, EQ-5D scores at baseline) were not recorded in the database, even though they could influence PROM responses and thus ultimately impact the case-mix adjustment model. This paper suggests that the specific predictors for the EQ-5D, based on this data, have not been found yet.

CONCLUSIONS

In conclusion, this study shows that other predictors (e.g. psychological and social factors) should be considered as potential case-mix variables for patient-reported outcome measures (PROMs) than for clinical outcomes in ischemic stroke patients. It is important that these specific case-mix variables should be included in order to benchmark hospitals legitimately on PROMs. One of the principles of value-based healthcare is to benchmark clinical outcomes and PROMs across different diseases and healthcare providers/ institutions to ensure quality improvement and competition.[36] This study identified a low(er) socio-economic status to be specifically associated with lower EQ-5D index scores. Future research should focus on finding other important predictors specific to PROMs in acute ischemic stroke to be able to further develop valid case-mix models.

AVAILABILITY OF DATA AND MATERIALS

The data that support the findings of this study are available from Hester Lingsma, PhD but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Hester Lingsma, PhD.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Formal ethical approval was not required as primary data were already collected and part of a quality registry. An 'opt-out' procedure was put in place, in which patients were informed about the quality registry and had the option to refuse their data being used.

COMPETING INTERESTS

- Authors AO and JH are supported by a grant from the Federation of Dutch University Medical Centers (NFU).
- Author DD reports grants and other from Stryker European Operations BV, grants from Penumbra Inc., grants from Medtronic, grants from Dutch Heart Foundation, grants from Brain Foundation Netherlands, grants from ZonMw and grants from Thrombolytic Science, LLC, outside the submitted work.
- Remaining authors declare no competing interests.
- This study was a cooperation of 8 parties: Stroke Knowledge Network Netherlands, Achmea Holding N.V., Netherlands Federation of University Medical Centres (NFU), Dutch Association for Neurology (NVN), Netherlands Society of Rehabilitation Medicine (VRA), Dutch Association of Elderly Care Physicians and Social Geriatricians (Verenso), The Dutch College of General Practitioners (NHG), Dutch Institute for Clinical Auditing (DICA), National Health Care Institute (ZN).

REFERENCES

1. Lovaglio PG. Benchmarking strategies for measuring the quality of healthcare: problems and prospects. *ScientificWorldJournal.* 2012.
2. Kuhrij LS, Wouters MW, van den Berg-Vos RM, de Leeuw FE, Nederkoorn PJ. The Dutch Acute Stroke Audit: Benchmarking acute stroke care in the Netherlands. *Eur Stroke J*. 2018;3(4):361–8.
3. Fayanju OM, Mayo TL, Spinks TE, Lee S, Barcenas CH, Smith BD, et al. Valuebased breast Cancer care: a multidisciplinary approach for defining patientcentered outcomes. *Ann Surg Oncol*. 2016;23(8):2385–90.
4. Katzan IL, Spertus J, Bettger JP, Bravata DM, Reeves MJ, Smith EE, et al. Risk adjustment of ischemic stroke outcomes for comparing hospital performance: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2014;45(3):918–44.
5. Kotronoulas G, Kearney N, Maguire R, Harrow A, Di Domenico D, Croy S, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. *J Clin Oncol*. 2014;32(14):1480–501.
6. Bredart A, Marrel A, Abetz-Webb L, Lasch K, Acquadro C. Interviewing to develop patient-reported outcome (PRO) measures for clinical research: eliciting patients' experience. *Health Qual Life Outcomes*. 2014;12:15.
7. Cella D, Lai JS, Nowinski CJ, Victorson D, Peterman A, Miller D, et al. NeureoQOL. Brief measures of health-related quality of life for clinical research in neurology. *Neurology*. 2012;78(23):1860–7.
8. EuroQol Group. EuroQol--a new facility for the measurement of healthrelated quality of life. *Health policy.* 1990;16(3):199–208.
9. Goldstein H, Spiegelhalter DJ. League tables and their limitations: statistical issues in comparisons of institutional performance. *J Royal Stat Society Series A (Statistics in Society).* 1996;159(3):385–443.
10. van Dishoeck AM, Lingsma HF, Mackenbach JP, Steyerberg EW. Random variation and rankability of hospitals using outcome indicators. *BMJ Qual Saf*. 2011;20(10):869–74.
11. Teale EA, Forster A, Munyombwe T, Young JB. A systematic review of casemix adjustment models for stroke. *Clin Rehabil.* 2012;26(9):771–86.
12. Appleby J, Devlin N, Parkin D. Using patient reported outcomes to improve health care. Chichester: Wiley; 2015.
13. Wilson JT, Hareendran A, Grant M, Baird T, Schulz UG, Muir KW, et al. Improving the assessment of outcomes in stroke: use of a structured interview to assign grades on the modified Rankin scale. *Stroke*. 2002;33(9):2243–6.
14. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff. *Ned Tijdschr Geneeskd*. 2005;149(28):1574–8.
15. van Agt HM, Essink-Bot ML, Krabbe PF, Bonsel GJ. Test-retest reliability of health state valuations collected with the EuroQol questionnaire. *Soc Sci Med*. 1994;39(11):1537–44.
16. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med Care*. 2005;43(3):203–20.
17. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997; 35(11):1095–108.

18. MV M, MV K, MAAE S, de Wit GA, Prenger R, AS E. Dutch tariff for the fivelevel version of EQ-5D. *Value Health*. 2016;19(4):343–52.
19. Cella D, Hahn EA, Jensen SE, Butt Z, Nowinski CJ, Rothrock N, et al. PatientReported Outcomes in Performance Measurement. Research Triangle Park (NC): RTI Press (c) 2015 Research Triangle Institute. All rights reserved. This book is protected by copyright. Credit must be provided to the author and source of the book when the content is quoted. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.; 2015.
20. Chen P, Lin K-C, Liing R-J, Wu C-Y, Chen C-L, Chang K-C. Validity, responsiveness, and minimal clinically important difference of EQ-5D-5L in stroke patients undergoing rehabilitation. *Qual Life Res*. 2016;25(6):1585–96.
21. Steyerberg EW. Clinical prediction models : a practical approach to development, validation and updating. New York: Springer; 2010.
22. Schaake K, Burgers J, Mulder CH. Ethnicity at the individual and neighborhood level as an explanation for moving out of the neighborhood. *Popul Res Policy Rev*. 2010;29(4):593–608.
23. Taphoorn MJ, Sizoo EM, Bottomley A. Review on quality of life issues in patients with primary brain tumors. *Oncologist*. 2010;15(6):618–26.
24. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Parzen E, Tanabe K, Kitagawa G, editors. Selected papers of Hirotugu Akaike. New York, NY: Springer New York; 1998. p. 199–213.
25. Colin Cameron A, Windmeijer FAG. An R-squared measure of goodness of fit for some common nonlinear regression models. *J Econ*. 1997; 77(2):329–42.
26. Bray BD, Campbell J, Cloud GC, Hoffman A, James M, Tyrrell PJ, et al. Derivation and external validation of a case mix model for the standardized reporting of 30-day stroke mortality rates. *Stroke*. 2014;45(11):3374–80.
27. Teale E, Young J, Dennis M, Sheldon T. Predicting patient-reported stroke outcomes: a validation of the six simple variable prognostic model. *Cerebrovasc Dis Extra*. 2013;3(1):97–102.
28. Wang R, Langhammer B. Predictors of quality of life for chronic stroke survivors in relation to cultural differences: a literature review. *Scand J Caring Sci*. 2018;32(2):502–14.
29. Carod-Artal FJ, Egido JA. Quality of Life after Stroke: The Importance of a Good Recovery. *Cerebrovasc Dis*. 2009;27(suppl 1):204–14.
30. Mar J, Masjuan J, Oliva-Moreno J, Gonzalez-Rojas N, Becerra V, Casado MÁ, et al. Outcomes measured by mortality rates, quality of life and degree of autonomy in the first year in stroke units in Spain. *Health Qual Life Outcomes*. 2015;13:36.
31. Reyes-Gibby CC, Anderson KO, Morrow PK, Shete S, Hassan S. Depressive symptoms and health-related quality of life in breast cancer survivors. *J Women's Health (Larchmt)*. 2012;21(3):311–8.
32. Wenzel L, DeAlba I, Habbal R, Kluhsman BC, Fairclough D, Krebs LU, et al. Quality of life in long-term cervical cancer survivors. *Gynecol Oncol*. 2005; 97(2):310–7.
33. Toni D, Lorenzano S, Puca E, Prencipe M. The SITS-MOST registry. *Neurol Sci*. 2006;27(3):s260–s2.
34. Sentinel Stroke National Audit Programme (SSNAP). https://www.rcplondon.ac.uk/projects/ outputs/sentinel-stroke-national-audit-programme-ssnap. Accessed September 24, 2019.

35. Bernhardt J, Zorowitz RD, Becker KJ, Keller E, Saposnik G, Strbian D, et al. Advances in stroke 2017. *Stroke*. 2018;49(5):e174–e99.
36. Porter ME, Teisberg EO. Redefining health care: creating value-based competition on results: *Harvard business press*; 2006.

Chapter 4

Case-Mix Adjustment Models for Hospital Comparisons in EQ-5D Domains after Ischemic Stroke: A Substudy of the MR CLEAN Trial

Arvind Oemrawsingh, MD MHS[1], Nikki van Leeuwen, PhD[1], Aad van der Lugt, MD, PhD[2]; Niek S. Klazinga, MD PhD[3]; Jan A. Hazelzet, MD PhD[1]; Jan J. van Busschbach, PhD[4], Diederik W.J. Dippel, MD PhD[5], Hester F. Lingsma, PhD[1]; on behalf of MR CLEAN investigators

[1] *Center for Medical Decision Making, Department of Public Health, Erasmus University Medical Center Rotterdam, PO Box 2040, 3000 CA, Rotterdam, the Netherlands*
[2] *Department of Radiology & Nuclear Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands*
[3] *Department of Public Health, Academic Medical Centre, University of Amsterdam, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands*
[4] *Department of Psychiatry, Section of Medical Psychology & Psychotherapy, Erasmus University Medical Center, Rotterdam, the Netherlands*
[5] *Department of Neurology, Erasmus University Medical Center, Rotterdam, the Netherlands*

Submitted.

ABSTRACT

Introduction

Patient-reported outcome measures (PROMs) may be applied in benchmarking hospitals. The aim of this study was to develop case-mix adjustment models for individual EQ-5D quality-of-life domains after acute ischemic stroke and to compare hospital effect estimates for domain scores with and without case-mix adjustment.

Patients & Methods

A secondary analysis was performed using the MR CLEAN randomized control trial data (2010-2015) of patients diagnosed with acute ischemic stroke in 12 stroke intervention centers. EQ-5D, a generic PROM, was measured at 90 days after inclusion. Multivariable case-mix models for EQ-5D domains were developed using ordinal regression models. Model performances were assessed with R2. Unadjusted and adjusted random effect estimates were compared to study the influence of case-mix adjustment on the variation in EQ-5D scores between centers.

Results

336 patients were analyzed. Common predictors in the regression models across all individual EQ-5D domains were hypertension and peripheral arterial occlusive disease. More predictors were identified for "self-care" (12 predictors), "mobility" (11 predictors) and "usual activities" (9 predictors) than for "pain/discomfort" (6 predictors) and "anxiety/depression" (4 predictors). R2 values ranged from 0.08 ("anxiety/depression"), 0.13 ("pain/discomfort"), 0.19 ("usual activities"), 0.27 ("mobility") to 0.31 ("self-care"). After case-mix adjustment, estimated performance of individual stroke centers changed only slightly for "physical" domains as opposed to "psychological" domains and the composite EQ-5D.

Discussion & Conclusion

Case-mix models developed for comparing quality of life scores after stroke between hospitals showed weak to modest predictive power for "anxiety/depression" and "pain/discomfort". These outcome domains might be less influenced by traditional (clinical) baseline characteristics, which implies a need to collect other variables for case-mix adjustment such as baseline PROMs, socio-economic status, education level, and specific comorbidities, when benchmarking PROMs.

INTRODUCTION

Patient-Reported Outcome Measures (PROMs) are increasingly being advocated as part of patient-centered and value-based healthcare (VBHC), and may be instrumental in quality of care improvement.[1] PROMs can be defined as patients' direct feedback on their health status and/or treatment, free from external interpretation by healthcare professionals.[2] A variety of multidimensional generic (e.g. EQ-5D, SF-36 (Short Form-36)) and disease-specific (e.g. SS-QOL (Stroke-Specific Quality of Life), SIS (Stroke Impact Scale)) PROMs are available for stroke patients.[3] Although routine collection of PRO(M)s has been recommended by the American Stroke Association[4], its implementation into clinical stroke care to improve quality of care is not yet widespread[3].

Traditionally, healthcare quality indicators were mainly clinical outcomes (e.g. mortality, complications of treatment) and these still remain key outcome measures. Variation in clinical outcomes between centers is largely explained by differences in case-mix[5], a phenomenon which has not been explored adequately for patient-reported outcome measures. Case-mix adjustment should capture differences between hospitals' patient characteristics that are outside the influence of the hospital, and/ but that influence the outcome of interest.[6] It is vital to correct for these patient characteristics, because outcome differences between hospitals may otherwise reflect differences in baseline risk.[7] If comparisons between hospitals are attempted, it is important to consider both the case-mix factor's predictive relationship with the outcome of interest as well as its distribution across hospitals.[8] The development of case-mix risk adjustment models has been previously reported for clinical outcomes in acute stroke[9], and it has been demonstrated that inadequate case-mix adjustment limits the validity of between-hospital comparisons[10].

PROMs have the advantage of providing additional information to healthcare professionals that can help them to make decisions for individual patients.[11] Routinely collected PROMs may also be potentially utilized for benchmarking the performance of health care services.[12,13] Despite a strong desire to use PROMs for benchmarking purposes[14], there still remain significant gaps in the evidence-base on the impact of routinely collected PROMs on the quality improvement of care providers[15,16]. In addition, there is also a scarcity of experience and knowledge on the development of case-mix adjustment models for PROMs. As the routine collection of PROMs is increasingly more common, the need for methodologically robust case-mix adjustment of PROM data for benchmarking purposes is essential.[17]

The aim of this study was to build case-mix adjustment models for the individual EQ-5D domains (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) and to compare hospital effect estimates for EQ-5D domain scores before and after case-mix adjustment, based on the MR CLEAN trial data.

METHODS

Anonymized trial data and syntax of statistical analyses that support the study findings are available from the principal investigator (e-mail: mrclean@erasmusmc.nl) on reasonable request.

Study Population

The Multicenter Randomized Clinical trial of Endovascular treatment for Acute ischemic stroke in the Netherlands (MR CLEAN) study[18,19] is a randomized clinical trial (RCT) comparing EVT (endovascular treatment with alteplase or urokinase, mechanical treatment or both) to usual care alone in patients with a radiologically confirmed proximal intracranial arterial occlusion, treatable within 6 hours after the onset of symptoms. This RCT was conducted in 16 Dutch stroke centers and originally included 500 patients. Patient eligibility, methods and study protocol of the MR CLEAN (2010 – 2015) trial have been previously reported.[18,19] Written informed consent was obtained before randomization from all patients or their legal representatives. The study protocol was approved from a central medical ethics committee and the research board of each participating center.

Figure 1 illustrates the selection of the original study population which was used for this substudy. Candidate model predictors that were missing in the original database (N = 500 patients), were imputed 10 times. Afterwards, four stroke centers that included less than 5 patients (N = 13), patients who died before the three-month follow-up (N = 108), and patients with missing EQ-5D data (N = 43) were excluded. Ultimately, after imputation, all analyses were performed on the remaining 336 patients treated in 12 stroke intervention centers.

Patient-Reported Outcome Measure

The EQ-5D-3L[20,21] was the patient-reported outcome measure used as a secondary outcome in the MR CLEAN trial. This generic health-related quality of life instrument has been previously translated in Dutch and validated.[22,23] The EQ-5D consists of five domains: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each domain has three levels of severity: 1 = no problems; 2 = some

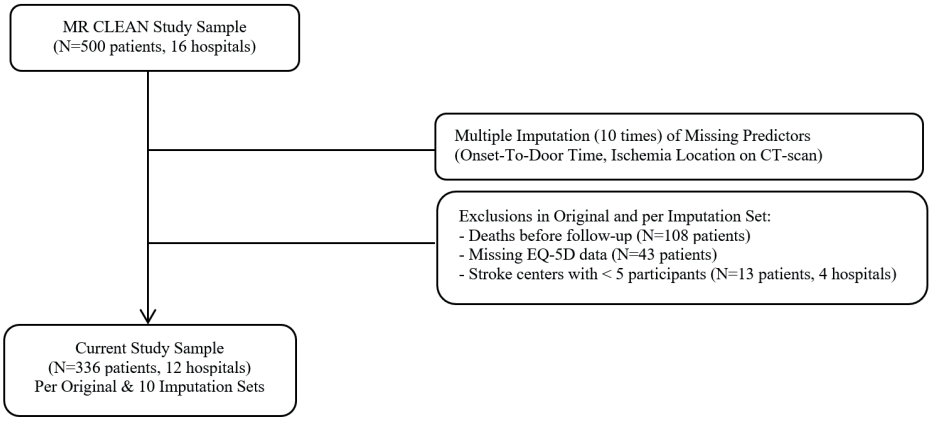


Figure 1 | Flowchart of Study Sample Selection for Case-Mix Models Development & Stroke Center Comparisons.

problems; 3 = extreme problems. The self-report questionnaire was telephone-administered to study participants or their proxy by an independent investigator at 90 days following inclusion.

In this substudy, the severity responses to each EQ-5D domain were used as outcomes for the individual ordinal regression models. The composite EQ-5D score, derived through the Dutch EQ-5D tariff[112], was used as an outcome for the linear regression model. The predicted values for the Dutch EQ-5D scores are anchored at −0.446 (state worse than death) and 1.00 (perfect health).[24] Scores can range from negative to positive values, with higher values signifying a better quality of life.

Case-Mix Risk Adjustment Models

Case-mix risk adjustment models for the five *individual* and *composite* EQ-5D domains were developed by initially putting all baseline patient characteristics as candidate predictors in univariable ordinal regression models for each individual EQ-5D domain as outcomes, while a linear regression model was developed for the composite EQ-5D score. These patient characteristics included gender, age, pre-stroke mRS (modified Rankin Scale) score, stroke onset-to-door time, location of intracranial arterial occlusion on vessel imaging (CTA, MRA or DSA), NIHSS (National Institute Health Stroke Scale) score at baseline, smoking and comorbidities (hypertension, diabetes, atrial fibrillation, peripheral arterial occlusive disease, previous stroke, hypercholesterolemia). The Akaike Information Criterion[25] (AIC; equivalent to $p < 0.157$), preferable for predictive models, was used as a cut-off to select predictors from the univariable regression models per outcome (five individual EQ-5D domains and the composite EQ-5D score). Afterwards, the

selected predictors were combined in multivariable regression models for each aforementioned outcome. Each model was completed by adding "EQ-5D report by proxy" as an explanatory variable in order to account for the variation in EQ-5D scores explained by the proxy's response as it has been reported that proxy respondents tend to overestimate impairments compared to patient self-reports.[26] In essence, all regression models were adjusted for the inter-rater gap, which represents the difference between patient-reported HRQoL (health-related quality of life) and the proxy ability to comprehend the patient's perspective.[27] Odds ratios (ORs) with corresponding p-values were calculated for the predictors in all final case-mix models.

Statistical Analysis

Descriptive statistics were presented as percentages or as a median ± inter-quartile range (IQR). The Pearson chi-squared statistic (categorical variables) and the Kruskal-Wallis test (continuous variables) were used to determine the unadjusted differences between the stroke centers. The significance level was set at p-value < 0.05.

To assess a stroke center's performance with respect to the EQ-5D composite and individual domain scores, hospitals were included as a categorical variable in both unadjusted and adjusted random-effects regression models per outcome. From both models, the βeta coefficients for each hospital were estimated and compared through a ratio in order to assess the impact of case-mix factors on hospital effect sizes. In addition, the relative difference in effect estimates before and after case-adjustment was calculated in percentages. The explained variance ('goodness-of-fit') of the regression models was assessed using R^2 statistics. Statistical analysis was carried out by using IBM SPSS Statistics 21 and RStudio version 1.0.153.0-© 2009-2016 Inc. software.

RESULTS

Study Population

In total, 336 patients from 12 participating hospitals were analyzed in the current substudy (Table 1). The number of included patients per stroke center ranged from 8 (hospital #7) to 56 (hospital #2). The median age was 63 years [IQR 52–73] and the majority of patients were male (59.5%). 108 patients (21.6%) died during the study period. The three-month EQ-5D data is presented in Table 2.

Table 1 | Descriptive Statistics Per Participating Stroke Center of Current Study Population (N=336 Patients, N=12 Hospitals).

	All (N= 336)	#1 (N=55)	#2 (N=56)	#3 (N=39)	#4 (N=14)	#5 (N=12)	#6 (N=16)	#7 (N=8)	#8 (N=38)	#9 (N=35)	#10 (N=10)	#11 (N=37)	#12 (N=16)
Male, N (%)	200 (59.5)	33 (60)	38 (67.9)	24 (61.5)	5 (35.7)	8 (66.7)	6 (37.5)	6 (75)	23 (60.5)	22 (62.9)	6 (60)	22 (59.5)	7 (43.8)
Age, median [IQR]	63 [52–73]	56 [49– 67]	64 [52.5–73]	67 [54–73]	56.5 [43.5–63.5]	62.5 [61–74.5]	76 [61 – 82.5]	59 [43.5–70]	68 [56–78]	63 [50–70]	64 [50–71]	67 [54.5–73]	60.5 [50–68]
Smoking (current), N (%)	105 (31.3)	14 (25.5)	21 (37.5)	8 (20.5)	3 (21.4)	4 (33.3)	2 (12.5)	3 (37.5)	15 (39.5)	14 (40)	2 (20)	13 (35.1)	6 (37.5)
Comorbidities, N (%)													
Previous stroke	29 (8.6)	4 (7.3)	4 (7.1)	7 (17.9)	2 (14.3)	0	1 (6.3)	0	4 (10.5)	1 (2.9)	2 (20)	4 (10.8)	0
Myocardial infarction	45 (13.4)	6 (10.9)	10 (17.9)	7 (17.9)	0	1 (8.3)	4 (25)	0	7 (18.4)	5 (14.3)	0	3 (8.1)	2 (12.5)
Diabetes mellitus	34 (10.1)	7 (12.7)	2 (3.6)	1 (2.6)	2 (14.3)	1 (8.3)	5 (31.3)	0	4 (10.5)	6 (17.1)	1 (10)	4 (10.8)	1 (6.3)
Hypertension	136 (40.5)	18 (32.7)	30 (53.6)	16 (41)	5 (35.7)	5 (41.7)	6 (37.5)	2 (25)	18 (47.4)	16 (45.7)	3 (30)	14 (37.8)	3 (18.8)
Hypercholesterolemia	75 (22.3)	7 (12.7)	14 (25)	8 (20.5)	1 (7.1)	3 (25)	6 (37.5)	1 (12.5)	14 (36.8)	7 (20)	0	11 (29.7)	3 (18.8)
PAOD	11 (3.3)	2 (3.6)	4 (7.1)	0	0	0	2 (12.5)	0	1 (2.6)	1 (2.9)	0	1 (2.7)	0
Atrial fibrillation	89 (26.5)	14 (25.5)	14 (25)	12 (30.8)	3 (21.4)	4 (33.3)	6 (37.5)	0	15 (39.5)	5 (14.3)	5 (50)	6 (16.2)	5 (31.3)
Pre-stroke mRS													
0 – 2	327 (97.3)	53 (96.4)	56 (100)	38 (97.4)	14 (100)	12 (100)	15 (93.8)	8 (100)	35 (92.1)	33 (94.3)	10 (100)	37 (100)	16 (100)
3 – 5	9 (2.7)	2 (3.6)	0	1 (2.6)	0	0	1 (6.2)	0	3 (7.9)	2 (5.7)	0	0	0

Table 1 | Descriptive Statistics Per Participating Stroke Center of Current Study Population (N=336 Patients, N=12 Hospitals). (continued)

	All (N= 336)	#1 (N=55)	#2 (N=56)	#3 (N=39)	#4 (N=14)	#5 (N=12)	#6 (N=16)	#7 (N=8)	#8 (N=38)	#9 (N=35)	#10 (N=10)	#11 (N=37)	#12 (N=16)
Location intra-arterial occlusion on CTA, N (%)													
ICA	4 (1.2)	1 (1.8)	0	0	0	0	0	1 (12.5)	1 (2.6)	0	0	1 (2.7)	0
ICA-T	88 (26.2)	16 (29.1)	15 (26.8)	7 (17.9)	4 (28.6)	5 (41.7)	2 (12.5)	4 (50)	9 (23.7)	10 (28.6)	3 (30)	10 (27)	3 (18.8)
M1	215 (64)	32 (58.2)	40 (71.4)	27 (69.2)	10 (71.4)	7 (58.3)	8 (50)	2 (25)	23 (60.5)	24 (68.6)	5 (50)	24 (64.9)	13 (81.3)
M2	26 (7.7)	4 (7.3)	1 (1.8)	5 (12.8)	0	0	6 (37.5)	1 (12.5)	4 (10.5)	1 (2.9)	2 (20)	2 (5.4)	0
A1	3 (0.9)	2 (3.6)	0	0	0	0	0	0	1 (2.6)	0	0	0	0
Systolic Blood pressure*, median [IQR]	140 [128–156]	135 [129–149]	145 [130–159]	140 [125–150]	130 [120–144]	137 [114–158]	151 [131–171]	135 [116–148]	142 [130–159]	147 [133–163]	132 [117–141]	149 [132–165]	135 [116–152]
NIHSS score at baseline, median [IQR]	17 [14–21]	19 [16– 22]	18 [14–22]	16 [13–19]	15 [13–18]	17 [17–24]	19 [14–23]	17 [15–21]	19 [13–23]	17 [11–20]	21 [16–24]	16 [12–20]	17 [13–22]
Intra-arterial treatment, N (%)	151 (44.9)	24 (43.6)	29 (51.8)	22 (56.4)	8 (57.1)	6 (50)	8 (50)	5 (62.5)	10 (26.3)	16 (45.7)	3 (30)	13 (35.1)	7 (43.8)
Intravenous thrombolysis, N (%)	301 (89.6)	43 (78.2)	51 (91.1)	35 (89.7)	14 (100)	11 (91.7)	15 (93.8)	8 (100)	35 (92.1)	32 (91.4)	7 (70)	35 (94.6)	15 (93.8)

Note. IQR = inter-quartile range; PAOD = peripheral arterial occlusive disease; CTA = CT angiography; ICA = internal carotid artery; ICA-T = internal carotid artery terminus; M1/M2 = middle cerebral artery; A1 = anterior cerebral artery; NIHSS = National Institute Health Stroke Scale;

Table 2 | EQ-5D Domain Responses of Current Study Population (N=336 Patients, N=12 Hospitals).

	All (N= 336)	#1 (N=55)	#2 (N=56)	#3 (N=39)	#4 (N=14)	#5 (N=12)	#6 (N=16)	#7 (N=8)	#8 (N=38)	#9 (N=35)	#10 (N=10)	#11 (N=37)	#12 (N=16)
EQ-5D domains at 3 months, N (%)													
Mobility													
1	137 (40.8)	20 (36.4)	28 (50)	16 (41)	7 (50)	6 (50)	5 (31.3)	4 (50)	16 (42.1)	13 (37.1)	2 (20)	15 (40.5)	5 (31.3)
2	177 (52.7)	30 (54.5)	25 (44.6)	20 (51.3)	7 (50)	6 (50)	7 (43.8)	4 (50)	19 (50)	19 (54.3)	7 (70)	22 (59.5)	11 (68.8)
3	22 (6.5)	5 (9.1)	3 (5.4)	3 (7.7)	0	0	4 (25)	0	3 (7.9)	3 (8.6)	1 (10)	0	0
Self-Care													
1	149 (44.3)	21 (38.2)	27 (48.2)	15 (38.5)	6 (42.9)	7 (58.3)	6 (37.5)	5 (62.5)	20 (52.6)	16 (45.7)	3 (30)	18 (48.6)	5 (31.3)
2	110 (32.7)	23 (41.8)	21 (37.5)	12 (30.8)	3 (21.4)	3 (25)	4 (25)	1 (12.5)	9 (23.7)	9 (25.7)	3 (30)	15 (40.5)	7 (43.8)
3	77 (22.9)	11 (20)	8 (14.3)	12 (30.8)	5 (35.7)	2 (16.7)	6 (37.5)	2 (25)	9 (23.7)	10 (28.6)	4 (40)	4 (10.8)	4 (25)
Usual Activities													
1	127 (37.8)	18 (32.7)	26 (46.4)	10 (25.6)	6 (42.9)	4 (33.3)	7 (43.8)	3 (37.5)	16 (42.1)	11 (31.4)	3 (30)	19 (51.4)	4 (25)
2	131 (39)	21 (38.2)	17 (30.4)	17 (43.6)	5 (35.7)	5 (41.7)	4 (25)	3 (37.5)	14 (36.8)	16 (45.7)	4 (40)	15 (40.5)	10 (62.5)
3	78 (23.2)	16 (29.1)	13 (23.2)	12 (30.8)	3 (21.4)	3 (25)	5 (31.3)	2 (25)	8 (21.1)	8 (22.9)	3 (30)	3 (8.1)	2 (12.5)
Pain													
1	146 (43.5)	18 (32.7)	30 (53.6)	14 (35.9)	5 (35.7)	6 (50)	6 (37.5)	4 (50)	22 (57.9)	8 (22.9)	5 (50)	22 (59.5)	6 (37.5)
2	140 (41.7)	29 (52.7)	17 (30.4)	16 (41)	8 (57.1)	4 (33.3)	7 (43.8)	2 (25)	12 (31.6)	22 (62.9)	2 (20)	14 (37.8)	7 (43.8)
3	50 (14.9)	8 (14.5)	9 (16.1)	9 (23.1)	1 (7.1)	2 (16.7)	3 (18.8)	2 (25)	4 (10.5)	5 (14.3)	3 (30)	1 (2.7)	3 (18.8)
Mood													
1	186 (55.4)	29 (52.7)	35 (62.5)	21 (53.8)	7 (50)	7 (58.3)	9 (56.3)	3 (37.5)	23 (60.5)	21 (60)	4 (40)	21 (56.8)	6 (37.5)
2	124 (36.9)	20 (36.4)	16 (28.6)	15 (38.5)	6 (42.9)	5 (41.7)	5 (31.3)	4 (50)	14 (36.8)	10 (28.6)	4 (40)	15 (40.5)	10 (62.5)
3	26 (7.7)	6 (10.9)	5 (8.9)	3 (7.7)	1 (7.1)	0	2 (12.5)	1 (12.5)	1 (2.6)	4 (11.4)	2 (20)	1 (2.7)	0
EQ-5D score, median [IQR]	0.69 [0.31–0.84]	0.61 [0.27–0.81]	0.73 [0.31–1.00]	0.55 [0.24–0.77]	0.68 [0.35–0.86]	0.78 [0.33–0.97]	0.71 [0.12–0.95]	0.69 [0.18–0.83]	0.71 [0.35–0.89]	0.69 [0.25–0.81]	0.33 [0.13–0.86]	0.78 [0.57–0.95]	0.62 [0.29–0.77]
EQ-5D assessment by proxy, N (%)	178 (53.3)	32 (58.2)	31 (55.4)	28 (71.8)	6 (42.9)	5 (41.7)	9 (56.3)	4 (50)	15 (40.5)	20 (58.8)	5 (50)	15 (40.5)	8 (50)
Missing, N (%)	2 (0.6)												

Note. IQR = inter-quartile range

Case-Mix Risk Adjustment Models

The estimated coefficients (ORs plus corresponding SE's) of the five ordinal (individual domains) and one linear (composite EQ-5D) regression models are presented in Table 3. Common predictors across all individual EQ-5D domains were peripheral arterial occlusive disease and hypertension. The strongest predictors (based on lowest p-values in multivariable models) for the individual EQ-5D domains were male gender (OR's ranging from 0.56 for "mobility" to 0.69 for "pain/discomfort"), middle cerebral artery occlusion (ORs ranging from 0.52 for "self-care" to 0.72 for "pain/discomfort"), diabetes (ORs ranging from 1.78 for "self-care" to 2.08 for "mobility"), and peripheral arterial occlusive disease (ORs ranging from 1.52 for "pain/discomfort" to 5.75 for "mobility").

The number of predictors in the ordinal regression models varied across each EQ-5D domain (excluding "EQ-5D assessment by proxy"). The range of independent variables per model ranged from 12 predictors ("self-care"), 11 predictors ("mobility"), 9 predictors ("usual activities"), 6 predictors ("pain/discomfort") to 4 predictors ("anxiety/depression"). The linear regression model for the composite EQ-5D contained 10 predictors including "EQ-5D assessment by proxy". Each case-mix adjusted model per individual EQ-5D domain had a moderate overall performance with R^2-values ranging from 0.08 ("anxiety/depression"), 0.13 ("pain/discomfort"), 0.19 ("usual activities"), 0.27 ("mobility") to 0.31 ("self-care"). The case-mix adjustment model for the composite EQ-5D had an R^2 value of 0.23.

Stroke Center Effect Estimates

Figure 2 demonstrates the effect of case-mix adjustment on stroke center effect estimates from the random effects ordinal and linear regression models (ratio between unadjusted and adjusted estimates) based on individual EQ-5D domain and composite EQ-5D scores within 12 hospitals in patients with ischemic stroke. Although most hospitals only had modest changes in stroke center effect estimates for all individual EQ-5D domains, in five stroke centers the effect reduction (after case-mix adjustment) was greater than or equal to 50% for "anxiety/depression" (hospitals #3, #6, and #9) and "pain/discomfort" (hospitals #1, #4, #6, #7, and #12). Only three hospitals (#6, #7, and #12) had a ≥ 50% reduction in their effect estimates for "mobility". For both "self-care" and "usual activities", only stroke centers #4, #6, #7, and #12 had a ≥ 50% reduction in their effect estimates. Hospital #6 was the only stroke center that had a ≥ 50% reduction in its effect estimates for *all* individual EQ-5D domains. Conversely, hospitals #2, #5, #8, #10, and #11 had modest or no changes in their effect estimates after case-mix adjustment for *all* individual EQ-5D domains.

Table 3 | Case-Mix Risk Adjustment Models.

	Self-Care		Mobility		Usual activities		Pain/ discomfort		Anxiety/ depression		EQ-5D	
R2 for model performance	**0.31**		**0.27**		**0.19**		**0.13**		**0.08**		**0.23**	
	OR	**p-value**	**OR**	**p-value**	**OR**	**p-value**	**OR**	**p-value**	**OR**	**p-value**	**βeta**	**p-value**
Age	1.01	0.177	1.01	0.262	1.004	0.664					1×10^{-4}	0.777
Gender (Reference: Female)	0.57	0.014	0.55	0.014			0.68	0.077			−0.07	0.039
Smoking	0.68	0.123			0.65	0.071					−0.05	0.181
Pre-stroke mRS score (Reference: mRS = 0)	1.28	0.452	0.96	0.914					1.32	0.380	0.02	0.681
NIHSS at baseline	1.01	0.565	1.04	0.102	1.04	0.081	1.02	0.299			−0.01	0.050
Location of ischemia (Reference: ICA & A1)	0.52	0.008	0.58	0.038	0.54	0.011	0.72	0.166				
PAOD	1.75	0.411	5.80	0.015	1.47	0.552	1.35	0.628	1.52	0.511	0.15	0.162
Diabetes	1.78	0.151	2.36	0.041	2.00	0.066					0.06	0.370
Hypertension	1.27	0.352	1.33	0.276	1.05	0.844	1.48	0.080	1.27	0.307	0.07	0.081
Atrial fibrillation	1.37	0.233	1.22	0.472	1.18	0.502					0.02	0.693
Hypercholesterolemia	1.12	0.699										
Onset-to-door time			1.00	0.051								
Assessment by proxy	0.16	<0.001	0.23	<0.001	0.30	<0.001	0.34	<0.001	0.41	<0.001	−0.25	<0.001

Note. OR = odds ratio; SE = standard error; mRS = modified Rankin Scale; NIHSS = National Institute Health Stroke Scale; ICA = internal carotid artery; A1 = anterior cerebral artery; PAOD = peripheral arterial occlusive disease

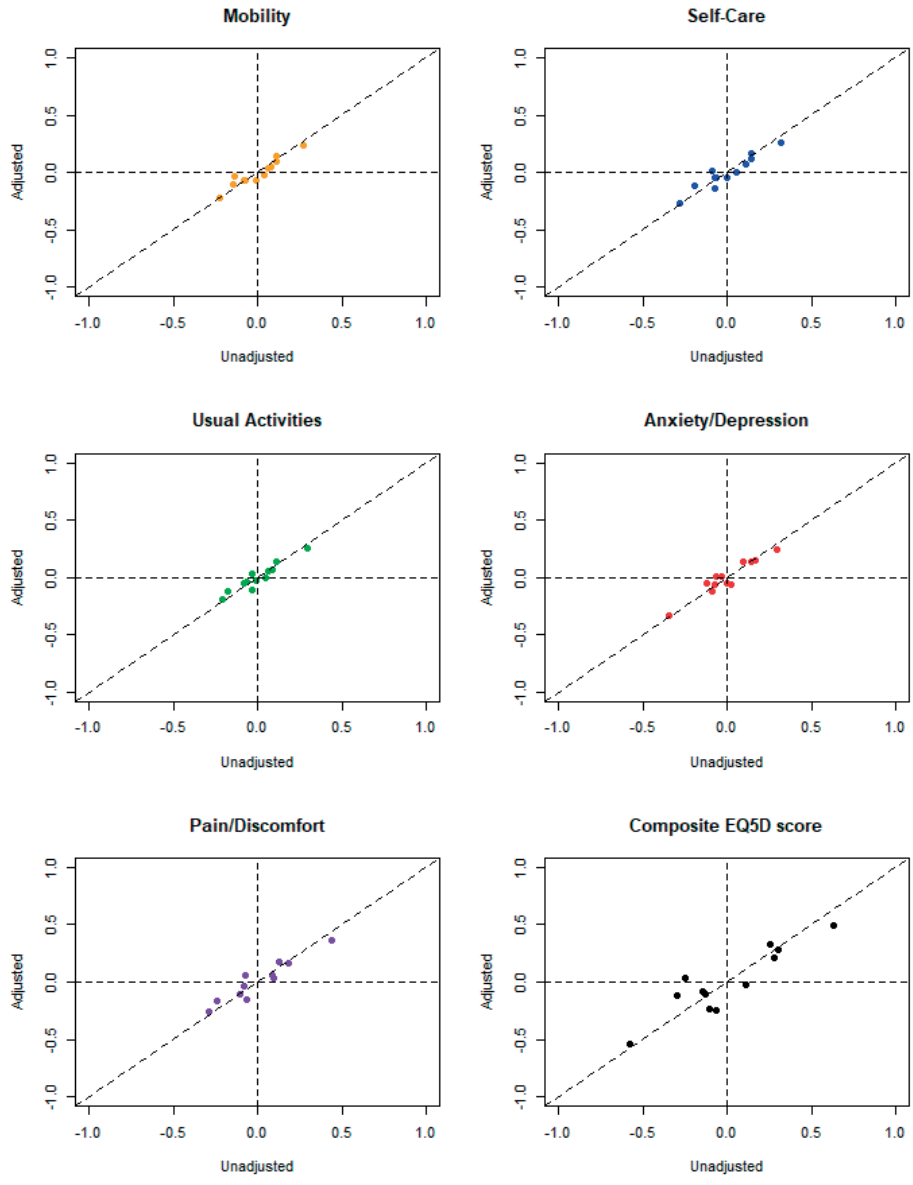


Figure 2 | Scatterplots for Unadjusted vs. Adjusted (with Random Effects) Stroke Center Estimates Per Individual EQ-5D Domain and Composite EQ-5D Scores.

Note.
- ● Each dot represents the ratio between adjusted/unadjusted effect estimates for a stroke center.
- – Dotted line is the line of equality (full agreement between unadjusted and adjusted estimates), and vertical distance away from this line is an indicator of the magnitude by which original unadjusted estimates have been reduced by adjusting for case-mix factors.

DISCUSSION

Payers and hospitals have been rapidly adopting PROMs, in addition to clinical and process indicators, as metrics to assess hospital performance in the context of value-based healthcare.[28] Although the demand for transparency in quality of care is increasing across hospitals, there still remains substantial knowledge gaps regarding the influence of random variation and case-mix factors in adequate hospital comparisons.[29] The aim of this substudy was to show the differences in case-mix adjustment models for the composite EQ-5D and its individual domains, and to determine the change in stroke center effect estimates after adjustment of case-mix factors, which are outside of the treating hospital's control. This study showed that case-mix adjustment models for "physical" EQ-5D domains ("mobility", "self-care" and "usual activities") contained more significant predictors and had better predictive power, with smaller differences between random unadjusted and adjusted stroke center effect estimates also being observed for "physical" domains.

To our knowledge, this is the first study that explores case-mix adjustment models for individual EQ-5D domains; therefore, comparison with previously published models is not possible. More significant case-mix predictors were identified in this study for the "physical" than "psychological" EQ-5D domains, with common predictors across all domains being peripheral arterial occlusive disease and hypertension. This study showed that male gender, age, baseline NIHSS score, smoking, hypertension and PAOD were strong predictors for the composite EQ-5D utility score at three months following trial inclusion. In a previously published methodological study on ischemic stroke patients from a stroke registry[30], similar associations were observed in a regression model with patient-reported EQ-5D scores as the outcome of interest. For the current substudy, the authors chose the Akaike's information criterion (AIC) as a pragmatic approach to model selection as the study population was moderate in size. It is generally known that the AIC will include more (weak) predictors as the sample size increases.[31]

This substudy also demonstrated that the difference between random unadjusted and adjusted stroke center effect estimates were larger for the composite EQ-5D and domain "pain/discomfort", as opposed to EQ-5D domains "mobility", "self-care" and "usual activities". These aforementioned results imply that the selected case-mix factors have a smaller impact on the "psychological" than the "physical" EQ-5D domains. As there was no data collected in the MR CLEAN trial on the occurrence of depression, it was not possible to corroborate the validity of

the case-mix model for "anxiety/depression". It is possible that other not included (psychological) case-mix variables such as pre-stroke anxiety/depression levels or socio-economic status have a bigger impact on psychological EQ-5D domain scores.[30] These observations coincide with the notion that many current case-mix adjustment models mainly consist of generic case-mix factors and lack disease-specific and psychological factors. Because the baseline EQ-5D status was not collected in the MR CLEAN trial, potential model adjustment by this possibly important predictor could unfortunately not be formally accounted for in this analysis. Capturing PROMs at different time points, especially prior to treatment (baseline), is recommended though it may be especially challenging for acute diseases such as acute ischemic stroke. If case-mix adjustment models include baseline PRO scores, socio-economic status and relevant comorbidities, benchmarking can be optimized in order to ultimately help discover potential for improvement and learning across the health care institutions. By identifying leaders in a particular field and understanding their (care) processes, other organizations may emulate or enhance their own processes.[32]

The MR CLEAN trial did not observe a significant difference (adjusted difference 0.06; 95% CI -0.01 – 0.18) in EQ-5D scores at 90 days following patient inclusion between the intervention (intra-arterial treatment) and control (usual care) arms, after excluding patients who died before the 90-day quality-of-life assessment.[18] Schreuders et al.[33] reanalyzed these EQ-5D scores by also including those patients who died during follow-up. The exact same EQ-5D difference was observed, compared to the previously published RCT results, but significant with a smaller 95% confidence interval. In the updated analysis, the authors also demonstrated that there was no treatment effect for the domains "pain/discomfort" and "anxiety/depression". This observation along with the current study's results, on the large differences between unadjusted and adjusted stroke center effect estimates particularly for these two aforementioned "psychological" domains, suggest that the EQ-5D may be too coarse a measure to not only reflect quality-of-life issues pertaining to stroke patients but also as an outcome to benchmark the quality of stroke care.

Strengths & Limitations

The current study is among the first in investigating the underlying influence of case-mix factors on the variation in individual EQ-5D domain scores. The key strength of this study was the randomized nature of the data source, specifically, a large RCT of 500 patients with limited missing 90-day EQ-5D data. Another strength of this study was that Dutch-specific tariffs were used to estimate the composite

EQ-5D utility score for Dutch trial participants, yielding Dutch parameter estimates that were used as the continuous outcome for the case-mix adjustment model. A limitation of this substudy were the large differences in the number of stroke patients per participating hospital. The hospitals with the smallest numbers hinder the detection of quality-of-care differences because of statistical uncertainty related to the limited power. The effect estimates of each small hospital will then often be too extreme (in relation to the mean estimate) and shift towards the mean in the random-effects analysis. This was partially circumvented by excluding four hospitals with less than five participating patients. Another limitation was the potential survivorship bias by excluding all patients who died before the three-month follow-up. Although the authors initially wanted to include them to reflect the total loss of quality of life due to stroke, there was no validated method to handle missing individual EQ-5D domain responses from deceased patients as outcomes for ordinal regression models.

CONCLUSION

This study demonstrated that case-mix models, developed for comparing quality of life scores after stroke between hospitals, showed weak to modest predictive power specifically for “anxiety/depression” and “pain/discomfort”. Traditional (clinical) baseline patient characteristics may have less influence on PRO domains, suggesting a need to collect other variables for case-mix adjustment. If PROs are to perform well as measures of quality of care, we recommend to not only collect different case-mix factors, potentially associated with PROs, but also PROMs at different stages in the care process (including prior to treatment, if possible).

ACKNOWLEDGEMENTS

A complete overview of all MR CLEAN investigators is provided in the Supplemental Material.

REFERENCES

1. Øvretveit J, Zubkoff L, Nelson EC, Frampton S, Knudsen JL, Zimlichman E. Using patient-reported outcome measurement to improve patient care. International Journal for Quality in Health Care. 2017;29(6):874-9.
2. Bredart A, Marrel A, Abetz-Webb L, Lasch K, Acquadro C. Interviewing to develop Patient-Reported Outcome (PRO) measures for clinical research: eliciting patients' experience. Health Qual Life Outcomes. 2014;12:15.
3. Reeves M, Lisabeth L, Williams L, Katzan I, Kapral M, Deutsch A, et al. Patient-Reported Outcome Measures (PROMs) for Acute Stroke: Rationale, Methods and Future Directions. Stroke. 2018;49(6):1549-56.
4. Rumsfeld JS, Alexander KP, Goff DC, Jr., Graham MM, Ho PM, Masoudi FA, et al. Cardiovascular health: the importance of measuring patient-reported health status: a scientific statement from the American Heart Association. Circulation. 2013;127(22):2233-49.
5. Fischer C, Lingsma HF, van Leersum N, Tollenaar RAEM, Wouters MW, Steyerberg EW. Comparing colon cancer outcomes: The impact of low hospital case volume and case-mix adjustment. European Journal of Surgical Oncology (EJSO). 2015;41(8):1045-53.
6. Lingsma HF, Steyerberg EW, Eijkemans MJ, Dippel DW, Scholte Op Reimer WJ, Van Houwelingen HC, et al. Comparing and ranking hospitals based on outcome: results from The Netherlands Stroke Survey. Qjm. 2010;103(2):99-108.
7. Fischer C, Lingsma H, Hardwick R, Cromwell DA, Steyerberg E, Groene O. Risk adjustment models for short-term outcomes after surgical resection for oesophagogastric cancer. Br J Surg. 2016;103(1):105-16.
8. McNamee R. Regression modelling and other methods to control confounding. Occupational and Environmental Medicine. 2005;62(7):500.
9. Teale EA, Forster A, Munyombwe T, Young JB. A systematic review of case-mix adjustment models for stroke. Clin Rehabil. 2012;26(9):771-86.
10. Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. Lancet. 2004;363(9415):1147-54.
11. Field J, Holmes MM, Newell D. PROMs data: can it be used to make decisions for individual patients? A narrative review. Patient Relat Outcome Meas. 2019;10:233-41.
12. Franklin P, Chenok K, Lavalee D, Love R, Paxton L, Segal C, et al. Framework To Guide The Collection And Use Of Patient-Reported Outcome Measures In The Learning Healthcare System. EGEMS (Wash DC). 2017;5(1):17.
13. Peters M, Crocker H, Jenkinson C, Doll H, Fitzpatrick R. The routine collection of patient-reported outcome measures (PROMs) for long-term conditions in primary care: a cohort survey. BMJ open. 2014;4(2):e003968.
14. Appleby, J., Devlin, N., Parkin, D., 2015. Using Patient Reported Outcomes to Improve Health Care. Chichester: Wiley.
15. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. BMC Health Serv Res. 2013;13:211.
16. Howell D, Molloy S, Wilkinson K, Green E, Orchard K, Wang K, et al. Patient-reported outcomes in routine cancer clinical practice: a scoping review of use, impact on health

outcomes, and implementation factors. Annals of oncology : official journal of the European Society for Medical Oncology. 2015;26(9):1846-58.
17. Hutchings A, Neuburger J, van der Meulen J, Black N. Estimating recruitment rates for routine use of patient reported outcome measures and the impact on provider comparisons. BMC Health Serv Res. 2014;14:66.
18. Berkhemer OA, Fransen PS, Beumer D, van den Berg LA, Lingsma HF, Yoo AJ, et al. A randomized trial of intraarterial treatment for acute ischemic stroke. N Engl J Med. 2015;372(1):11-20.
19. Fransen PS, Beumer D, Berkhemer OA, van den Berg LA, Lingsma H, van der Lugt A, et al. MR CLEAN, a multicenter randomized clinical trial of endovascular treatment for acute ischemic stroke in the Netherlands: study protocol for a randomized controlled trial. Trials. 2014;15:343.
20. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: Past, Present and Future. Appl Health Econ Health Policy. 2017;15(2):127-37.
21. EuroQol G. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy. 1990;16(3):199-208.
22. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. [Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff] Kwaliteit van leven meten in economische evaluaties: het Nederlands EQ-5D-tarief. Ned Tijdschr Geneeskd. 2005;149(28):1574-8.
23. van Agt HM, Essink-Bot ML, Krabbe PF, Bonsel GJ. Test-retest reliability of health state valuations collected with the EuroQol questionnaire. Soc Sci Med. 1994;39(11):1537-44.
24. M MV, K MV, S MAAE, de Wit GA, Prenger R, E AS. Dutch Tariff for the Five-Level Version of EQ-5D. Value Health. 2016;19(4):343-52.
25. Atkinson AC. A Note on the Generalized Information Criterion for Choice of a Model. Biometrika. 1980;67(2):413-8.
26. Oczkowski C, O'Donnell M. Reliability of proxy respondents for patients with stroke: a systematic review. J Stroke Cerebrovasc Dis. 2010;19(5):410-6.
27. Pickard AS, Knight SJ. Proxy evaluation of health-related quality of life: a conceptual framework for understanding multiple proxy perspectives. Med Care. 2005;43(5):493-9.
28. Sheckter CC, Matros E, Lee GK, Selber JC, Offodile AC, 2nd. Applying a value-based care framework to post-mastectomy reconstruction. Breast Cancer Res Treat. 2019;175(3):547-51.
29. van Dishoeck AM, Looman CW, van der Wilden-van Lier EC, Mackenbach JP, Steyerberg EW. Displaying random variation in comparing hospital performance. BMJ Qual Saf. 2011;20(8):651-7.
30. Oemrawsingh A, van Leeuwen N, Venema E, Limburg M, de Leeuw FE, Wijffels MP, et al. Value-based healthcare in ischemic stroke care: case-mix adjustment models for clinical and patient-reported outcomes. BMC Med Res Methodol. 2019;19(1):229.
31. Heinze G, Wallisch C, Dunkler D. Variable selection - A review and recommendations for the practicing statistician. Biom J. 2018;60(3):431-49.
32. Thonon F, Watson J, Saghatchian M. Benchmarking facilities providing care: An international overview of initiatives. SAGE open medicine. 2015;3:2050312115601692.
33. Schreuders J, van den Berg LA, Fransen PSS, Berkhemer OA, Beumer D, Lingsma HF, et al. Quality of life after intra-arterial treatment for acute ischemic stroke in the MR CLEAN trial—Update. International Journal of Stroke. 2017;12(7):708-12.

SUPPLEMENTAL MATERIAL

Table: Modified Rankin Scale at 3 months (N = 336 patients, N = 12 hospitals)

	All (N= 336)	#1 (N=55)	#2 (N=56)	#3 (N=39)	#4 (N=14)	#5 (N=12)	#6 (N=16)	#7 (N=8)	#8 (N=38)	#9 (N=35)	#10 (N=10)	#11 (N=37)	#12 (N=16)
mRS at 3 months, N (%)													
6	6 (1.8)	0	1 (1.8)	0	0	0	2 (12.5)	1 (12.5)	0	1 (2.9)	0	0	1 (6.3)
5	32 (9.5)	3 (5.5)	7 (12.5)	3 (7.7)	1 (7.1)	0	1 (6.3)	1 (12.5)	5 (13.2)	1 (2.9)	1 (10)	8 (21.6)	1 (6.3)
4	76 (22.6)	11 (20)	12 (21.4)	9 (23.1)	3 (21.4)	4 (33.3)	1 (6.3)	1 (12.5)	10 (26.3)	10 (28.6)	3 (30)	8 (21.6)	4 (25)
3	76 (22.6)	15 (27.3)	14 (25)	10 (25.6)	3 (21.4)	1 (8.3)	4 (25)	4 (50)	5 (13.2)	6 (17.1)	0	10 (27)	4 (25)
2	110 (32.7)	20 (36.4)	17 (30.4)	11 (28.2)	6 (42.9)	7 (58.3)	5 (31.3)	1 (12.5)	12 (31.6)	13 (37.1)	3 (30)	9 (24.3)	6 (37.5)
1	36 (10.7)	6 (10.9)	5 (8.9)	6 (15.4)	1 (7.1)	0	3 (18.8)	0	6 (15.8)	4 (11.4)	3 (30)	2 (5.4)	0

MR CLEAN Investigators

Executive Committee

Diederik W.J. Dippel (Department of Neurology, Erasmus MC University Medical Center); Aad van der Lugt (Radiology, Erasmus MC University Medical Center); Charles B.L.M. Majoie (Department of Radiology, Academic Medical Center, Amsterdam); Yvo B.W.E.M. Roos (Neurology, Academic Medical Center, Amsterdam); Robert J. van Oostenbrugge (Department of Neurology, Maastricht University Medical Center and Cardiovascular Research Institute Maastricht (CARIM)); Wim H. van Zwam (Radiology, Maastricht University Medical Center and Cardiovascular Research Institute Maastricht (CARIM)); Olvert A. Berkhemer (Department of Neurology, Erasmus MC University Medical Center; Department of Radiology, Academic Medical Center, Amsterdam); Puck S.S. Fransen (Department of Neurology and Radiology, Erasmus MC University Medical Center); Debbie Beumer (Department of Neurology, Erasmus MC University Medical Center; Department of Neurology, Maastricht University Medical Center and Cardiovascular Research Institute Maastricht (CARIM), the Netherlands); Lucie A. van den Berg (Neurology, Academic Medical Center, Amsterdam)

Local Principal Investigators

Wouter J. Schonewille (Department of Neurology, Sint Antonius Hospital, Nieuwegein, the Netherlands); Jan Albert Vos (Radiology, Sint Antonius Hospital, Nieuwegein, the Netherlands); Charles B.L.M. Majoie (Department of Radiology, Academic Medical Center, Amsterdam, the Netherlands); Yvo B.W.E.M. Roos (Neurology, Academic Medical Center, Amsterdam, the Netherlands); Paul J. Nederkoorn (Neurology, Academic Medical Center, Amsterdam, the Netherlands); Marieke J.H. Wermer (Department of Neurology, Leiden University Medical Center, the Netherlands); Marianne A.A. van Walderveen (Radiology, Leiden University Medical Center, the Netherlands); Robert J. van Oostenbrugge (Department of Neurology, Maastricht University Medical Center and Cardiovascular Research

Institute Maastricht (CARIM), the Netherlands); Wim H. van Zwam (Radiology, Maastricht University Medical Center and Cardiovascular Research Institute Maastricht (CARIM), the Netherlands); Julie Staals (Department of Neurology, Maastricht University Medical Center and Cardiovascular Research Institute Maastricht (CARIM), the Netherlands); Jeannette Hofmeijer (Department of Neurology, Rijnstate Hospital, Arnhem, the Netherlands); Jacques A. van Oostayen (Radiology, Rijnstate Hospital, Arnhem, the Netherlands); Geert J. Lycklama à Nijeholt (Department of Radiology, Haaglanden MC, the Hague, the Netherlands); Jelis Boiten (Neurology, Haaglanden MC, the Hague, the Netherlands); Diederik W.J. Dippel (Department of Neurology, Erasmus MC University Medical Center, the Netherlands); Patrick A. Brouwer (Radiology, Erasmus MC University Medical Center, the Netherlands); Bart J. Emmer (Radiology, Erasmus MC University Medical Center, the Netherlands); Sebastiaan F. de Bruijn (Department of Neurology, HAGA Hospital, the Hague, the Netherlands); Lukas C. van Dijk (Radiology, HAGA Hospital, the Hague, the Netherlands); L. Jaap Kappelle (Department of Neurology, University Medical Center Utrecht, the Netherlands); Rob H. Lo (Radiology, University Medical Center Utrecht, the Netherlands); Ewoud J. van Dijk (Department of Neurology, Radboud University Medical Center, Nijmegen, the Netherlands); Joost de Vries (Neurosurgery, Radboud University Medical Center, Nijmegen, the Netherlands); Paul L.M. de Kort (Department of Neurology,Sint Elisabeth Hospital, Tilburg, the Netherlands); Jan S.P. van den Berg (Department of Neurology, Isala Klinieken, Zwolle, the Netherlands); Willem Jan J. van Rooij (Department of Neurology, Isala Klinieken, Zwolle, the Netherlands); Boudewijn A.A.M. van Hasselt (Radiology, Isala Klinieken, Zwolle, the Netherlands); Leo A.M. Aerden (Department of Neurology, Reinier de Graaf Gasthuis, Delft, the Netherlands); René J. Dallinga (Radiology,Reinier de Graaf Gasthuis, Delft, the Netherlands); Marieke C. Visser (Department of Neurology, VU Medical Center, Amsterdam, the Netherlands); Joseph C.J. Bot (Radiology, VUMedical Center, Amsterdam, the Netherlands); Patrick C. Vroomen (Department of Neurology, University Medical Center Groningen, the Netherlands); Omid Eshghi (Radiology, University Medical Center Groningen, the Netherlands); Tobien H.C.M.L. Schreuder (Department of Neurology, Atrium Medical Center, Heerlen, the Netherlands); Roel J.J.Heijboer (Radiology, Atrium Medical Center, Heerlen, the Netherlands); Koos Keizer (Department of Neurology, Catharina Hospital, Eindhoven, the Netherlands); Alexander V. Tielbeek (Radiology, Catharina Hospital, Eindhoven, the Netherlands); Heleen M. den Hertog (Department of Neurology, Medical Spectrum Twente, Enschede, the Netherlands); Dick G. Gerrits (Radiology, Medical Spectrum Twente, Enschede, the Netherlands); Renske M. van den Berg-Vos (Department of Neurology, Sint Lucas Andreas Hospital, Amsterdam, the Netherlands); Giorgos B. Karas (Radiology, Sint Lucas Andreas Hospital, Amsterdam, the Netherlands)

Imaging Assessment Committee

Charles B.L.M. Majoie (Chair, Department of Radiology, Academic Medical Center, Amsterdam, the Netherlands); Wim H. van Zwam (Radiology, Maastricht University Medical Center and Cardiovascular Research Institute Maastricht (CARIM), the Netherlands); Aad van der Lugt (Radiology, Erasmus MC University Medical Center, the Netherlands); Geert J. Lycklama à Nijeholt (Department of Radiology, Haaglanden MC, the Hague, the Netherlands); Marianne A.A. van Walderveen (Radiology, Leiden University Medical Center, the Netherlands); Joseph C.J. Bot (Radiology, VU Medical Center, Amsterdam, the Netherlands); Henk A. Marquering (Biomedical Engineering and Physics, Academic Medical Center, Amsterdam, the Netherlands); Ludo F. Beenen (Department of Radiology, Academic Medical Center, Amsterdam, the Netherlands); Marieke E.S. Sprengers (Department of Radiology, Academic Medical Center, Amsterdam, the Netherlands); Sjoerd F.M. Jenniskens (Radiology, Radboud University Medical Center, Nijmegen, the Netherlands); René van den Berg (Department of Radiology, Academic Medical Center, Amsterdam, the Netherlands); Olvert A. Berkhemer (Department of Neurology, Erasmus MC University Medical Center; Department of Radiology, Academic Medical Center, Amsterdam, the Netherlands); Albert J. Yoo (Department of Radiology, Texas Stroke Institute)

Outcome Assessment Committee

Yvo B.W.E.M. Roos (Chair, Neurology, Academic Medical Center, Amsterdam, the Netherlands); Peter J. Koudstaal (Department of Neurology, Erasmus MC University Medical Center, the Netherlands); Jelis Boiten (Department of Neurology, Haaglanden MC, the Hague, the Netherlands); Ewoud J. van Dijk (Department of Neurology, Radboud University Medical Center, Nijmegen, the Netherlands)

Adverse Event Committee

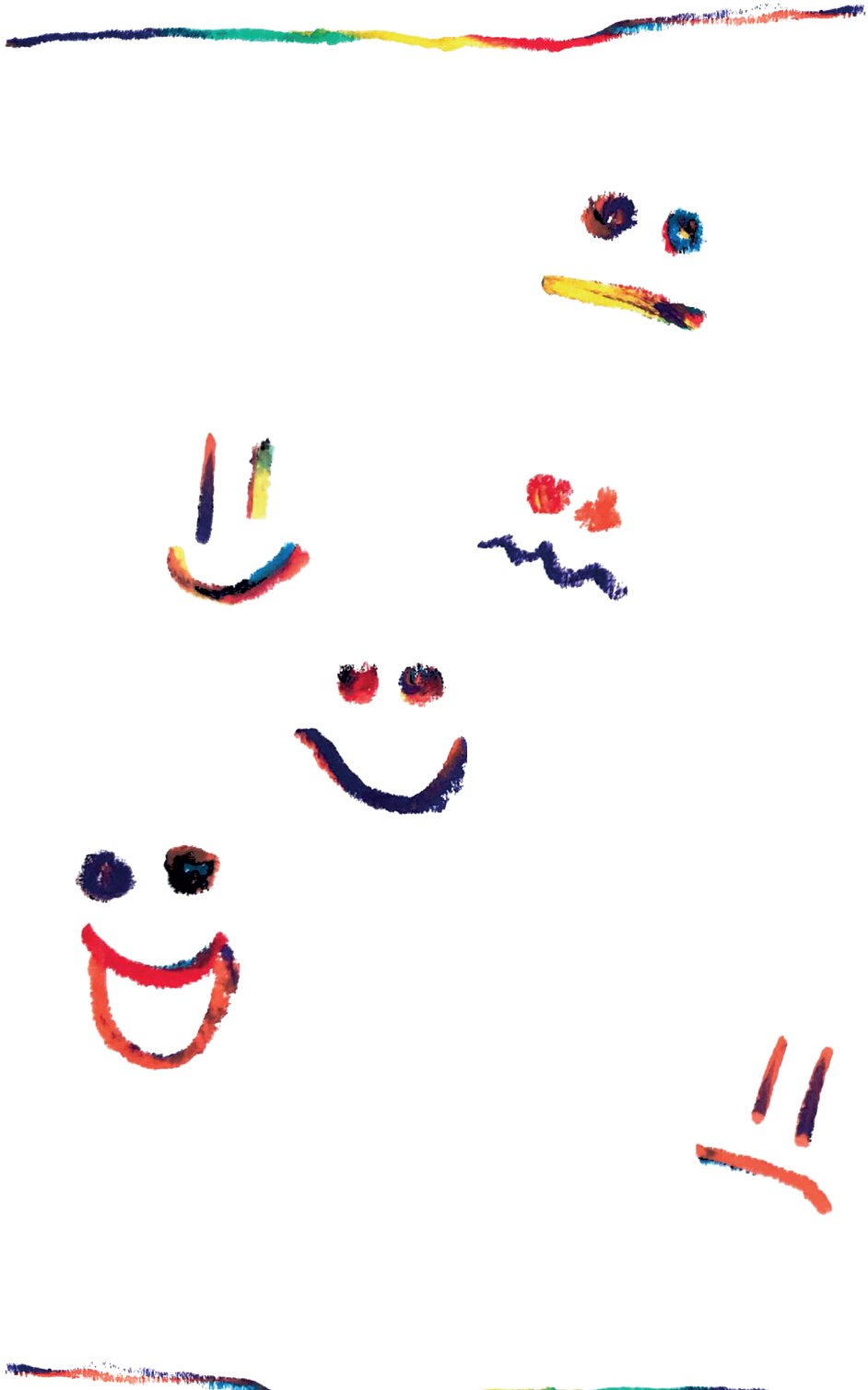
Robert J. van Oostenbrugge (Chair, Department of Neurology, Maastricht University Medical Center and Cardiovascular Research Institute Maastricht (CARIM), the Netherlands); Marieke J.H. Wermer (Department of Neurology, Leiden University Medical Center, the Netherlands); H. Zwenneke Flach (Radiology, Isala Klinieken, Zwolle, the Netherlands)

Trial Statisticians

Ewout W. Steyerberg (Public Health, Erasmus MC University Medical Center, the Netherlands; Medical Statistics and Bioinformatics, Leiden University Medical Center, the Netherlands); Hester F. Lingsma (Public Health, Erasmus MC University Medical Center, the Netherlands).

PART II

Practical Applications of PROMs in Clinical Care



Chapter 5 Implementing Patient-Reported Outcome Measures in Clinical Breast Cancer Care: A Systematic Review

Laurentine S.E. van Egdom, MD[1ǂ]; **Arvind Oemrawsingh, MD MHS[2ǂ]**, Lisanne M. de Jong-Verweij, PhD[2]; Hester F. Lingsma, PhD[2]; Linetta B. Koppert, MD PhD MHS[1]; Niek S. Klazinga, MD PhD[3]; Jan A. Hazelzet, MD PhD[2]

ǂ **Shared co-first authorship**

[1] *Department of Surgical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands*
[2] *Center for Medical Decision Making, Department of Public Health, Erasmus University Medical Center Rotterdam, PO Box 2040, 3000 CA, Rotterdam, the Netherlands*
[3] *Department of Public Health, Academic Medical Center, University of Amsterdam, Amsterdam Public Health Research Institute, Amsterdam, the Netherlands*

Value in Health, 2019 Oct;22(10):1197-1226.

ABSTRACT

Background

Patient-reported outcome measures (PROMs) are increasingly being used to improve care delivery and are becoming part of routine clinical practice.

Objective

This systematic review aims to give an overview of PROM administration methods and their facilitators and barriers in breast cancer clinical practice.

Methods

A systematic literature search was conducted in Embase, MEDLINE, PsycINFO, Cochrane Central, CINAHL, and Web of Science for potentially relevant articles from study inception to November 2017. Reference lists of screened reviews were also checked. After inclusion of relevant articles, data were extracted and appraised by 2 investigators.

Results

A total of 2311 articles were screened, of which 34 eligible articles were ultimately included. Method and frequency of PROM collection varied between studies. The majority of studies described a promising effect of PROM collection on patients (adherence, symptom distress, quality of life, acceptability, and satisfaction), providers (willingness to comply, clinical decision making, symptom management), and care process or system outcomes (referrals, patient-provider communication, hospital visits). A limited number of facilitators and barriers were identified, primarily of a technical and behavioral nature.

Conclusion

Although interpreting the impact of PROM collection in breast cancer care is challenging owing to considerations of synergistic (multicomponent) interventions and generalizability issues, this review found that systematic PROM collection has a promising impact on patients, providers, and care processes/ systems. Further standardization and reporting on method and frequency of PROM collection might help increase the effectiveness of PROM interventions and is warranted to enhance their overall impact.

Keywords

Breast Cancer Care; Implementation; Patient-Reported Outcome Measures; Value

INTRODUCTION

Breast cancer is the most common cancer affecting women worldwide.[1] With survival rates continuing to improve, the focus on quality of life is becoming increasingly important. Because healthcare is shifting toward a more value-based framework for quality of care improvement, more attention is being paid to patient-reported outcomes (PROs).[2] PROs are defined as feedback on a patient's health condition (ie, symptoms and quality of life) coming directly from the individual patient, thus without external interpretation.[3] Measurement of these outcomes is based on self-completed questionnaires called patient-reported outcome measures (PROMs).

PROMs are increasingly being collected and advocated in cancer care for aiding care management of the individual patient.[4, 5, 6, 7] The use of PROMs has been associated with better patient satisfaction,[8] perceptions of quality of care,[9, 10, 11] health outcomes,[12, 13, 14] and higher patient acceptability.[12, 15] Routine collection of PROMs can also have a positive impact on patient–provider communication, (shared) decision making, and symptom management.[8, 16, 17] Challenges in collecting, storing, analyzing, and reporting PROM scores in real time can impede the routine use of PROMs in clinical practice,[5] making it important to identify these implementation issues. Howell et al[17] published a review of the factors of PROM implementation and use in cancer clinical practice and concluded that PROMs have been tested most often in the breast cancer population. However, reviews focusing on methods of PROM administration, specifically in breast cancer care, have not been published.

This review sought to provide an overview of PROM collection methods in breast cancer care answering the following questions: (1) how have PROMs been administered in breast cancer care; (2) what is the impact of PROM administration on patients, care providers, and healthcare services or processes; and (3) what are the facilitators and barriers that influence the integration of PROM collection in routine breast cancer clinical practice?

METHODS

This systematic literature review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses model[18] for reporting of systematic reviews.

Literature Search Strategy

Six electronic databases were searched—Embase, MEDLINE, PsycINFO, Cochrane Central, CINAHL and Web of Science—from study inception to November 3, 2017. No restrictions to language or country of publication were applied to either the search strategy or study selection. The comprehensive search strategy was devised jointly with an experienced librarian and included a combination of keywords for breast cancer, which were searched in free text and as exploded medical subject headings where possible. Additional related terms were used to maximize the sensitivity of the search. The full search strategy is provided in the Supplemental Material.

Study Selection

Figure 1 shows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of the study selection process. Two investigators (A.O. and L.E.) independently screened titles and abstracts and identified potentially relevant articles.

The reference lists of identified reviews were screened for other relevant publications. Studies were excluded if they (1) did not include breast cancer patients, (2) had an irrelevant aim by focusing on PROMs as an evaluation method, endpoint, or outcome for other interventions and treatments, (3) had the study design of a case report, editorial, review, or study protocol, (4) had an overlapping study population with a related study (if both were relevant, the most recent article was included), and (5) were not published in full text.

Remaining studies were reviewed in full text by both authors using the same exclusion criteria. Disagreements on study eligibility were resolved through discussion between both assessors or, ultimately, through consultation with the whole research team. Subsequently included full-text publications were listed in a taxonomy table that comprised study descriptives (number of involved patients and healthcare providers, method and frequency of PROM collection, outcomes of PROM collection, and facilitators or barriers).

Data Extraction and Quality Assessment

Data that were extracted included information on study aim and design, sample characteristics (for intervention and control groups, if applicable), and method or frequency of PROM collection. Outcomes described in the included studies were divided into 3 categories during data extraction: (1) patient outcomes (ie, adherence, symptoms, health outcomes, acceptability, and satisfaction), (2) care provider outcomes (ie, adherence, impact on clinical decision making, acceptability,

and satisfaction), and (3) care process or system outcomes (ie, impact on referral [rates], communication, hospital visits, and usability). Acceptability was defined as the extent to which PRO collection was found pleasant by study participants. Satisfaction was defined as the extent to which the study participants enjoyed completing the PRO instruments.[19]

Finally, facilitators and barriers of routine PROM use in breast cancer care that were described were also extracted. The Critical Appraisal Skills Programme tool was used by both reviewers (A.O. and L.E.) to evaluate the quality of included studies (see Table in Supplementary Materials found at https://doi-org.eur.idm.oclc.org/10.1016/j-jval.2019.04.1927). Discrepancies were discussed by both reviewers and supervisors until consensus was reached. There was no meta-analysis performed owing to the heterogeneity among the included studies.

RESULTS

After the initial screening of titles and abstracts (n = 2311), 58 potentially relevant articles were retrieved (Figure 1). These articles were checked for eligibility based on full-text assessment. Thirty-four studies ultimately met the inclusion criteria (Table 1).

Total Study Population

The total sum of patients across all selected studies was 14 083, of which 11 191 (79.5%) were breast cancer patients (range 13-8359 patients). Sixteen of the 34 included articles exclusively involved breast cancer patients. In the remaining 18 studies, breast cancer patients were a subgroup of study populations with various cancer types (prostate, lung, colon, gynecologic, and genitourinary cancer, among others). Mean age was 54.68 years (± standard deviation [SD] 5.01).

Implementation

The patient-reported outcome version of the Common Terminology Criteria of Adverse Events[19-24, 35] was used most often, followed by the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire module QLQ-C30[25-27, 29, 30, 49], the EQ-5D questionnaire[24, 27, 35, 36], the M. D. Anderson Symptom Inventory[31-33], the Functional Assessment of Cancer Therapy questionnaires[31, 33, 37], the Hospital Anxiety and Depression Scale[26, 30, 39], and the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire moduleQLQ-B23[11, 25], respectively.

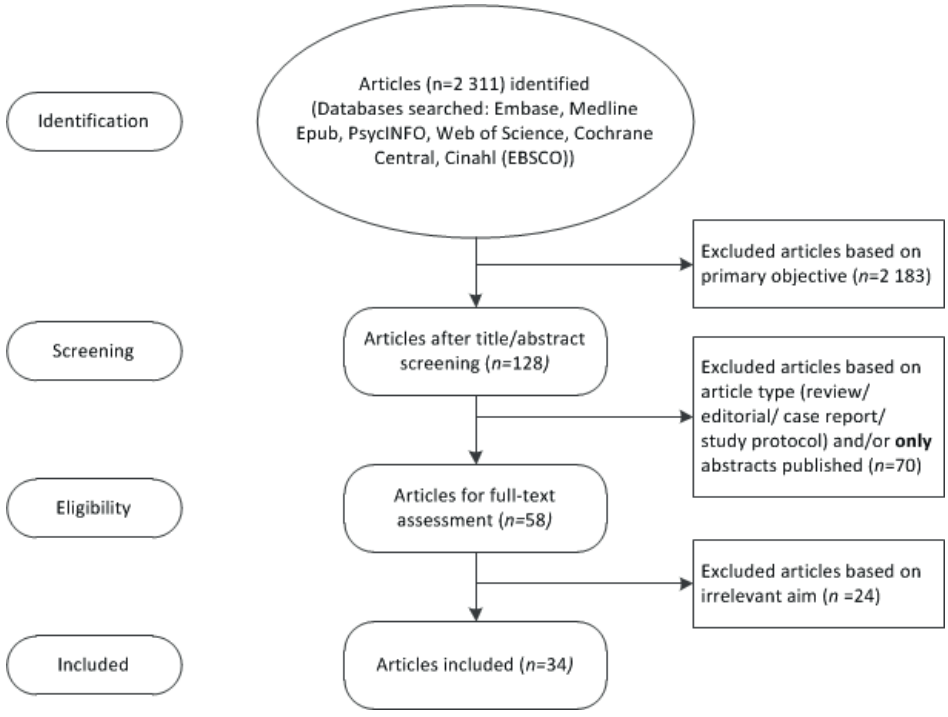


Figure 1 | Preferred Reporting Items for Systematic Reviews and Meta-Analyses Flow Diagram of Relevant Article Selection.

PROMs were most often collected electronically (n = 17). Web-based assessments were the primary method of data collection (n = 7), both in and outside the clinic, followed by the use of a tablet (n = 6), an (e-Health) application (n = 2), email (n = 1), and a software system (n = 1) (Table 1). Four studies[22, 31, 33, 40] used both electronic and paper-based interventions to collect PROMs. A telephone-based intervention was used in 9 studies, with 4 studies involving interactive voice response[32, 41, 42, 43], 3 studies involving mobile applications[20, 44, 45], and 2 studies involving semi-structured (computer-assisted) telephone interviews or structured prompts of PRO domains.[26, 28] One study used only a paper-based intervention.[27] Two qualitative studies involving patient focus groups did not include a PROM intervention.[46, 47] One study[48] did not specify the way PRO scores were collected.

The frequency of PROM administration varied from once to daily during the various study periods (Table 1). The duration of PROM interventions ranged from 1.5 months[20] to 24 months.[48] Most interventions (n = 14) provided PROMs only in the clinic, while 8 studies required patients to complete PROMs at home (all electronic- or telephone-based interventions), and 3 studies allowed patients to

complete PROMs either at home or the clinic. PROMs were collected during treatment (n = 22), during follow-up (n = 3), or during both (n = 9). One third of PROM interventions restricted assessments to monitor only a specific phase of breast cancer treatment, most commonly during chemotherapy. The most common methods for reminders were email or phone.

In 4 studies,[11, 21, 29, 43] the electronic PRO systems were directly integrated into the electronic health record. Some PROM interventions also provided patient education (n = 9); one of these[35] was administered through a module. Of 10 studies[11, 19, 22, 23, 25, 26, 28, 29, 46, 47] describing healthcare provider involvement (ie, focus group, interviews, satisfaction or usability questionnaires), only 2 studies[19, 22] described providers actually being part of the PROM intervention. Within the PROM intervention, patients' providers could edit automatically generated care plans (based on PROM responses) to further tailor referrals and treatment recommendations. Once approved by providers, care plans were mailed to patients.[19, 22] Outcomes were discussed by the treating clinician (n = 9), nurse (practitioner) or physician assistant (n = 5), or by a combination of these (n = 7) (Table 1). Nearly one third of PROM interventions included alerts being sent to clinicians or patients. Alerts were typically sent by email or text message. In 18 studies, summary reports of PRO data were sent to prespecified providers.

Impact of PROMs on Patient Outcomes

Adherence

Among studies reporting on adherence (n = 14), completion rates ranged between 71%[32] and 100%[48] (Table 1). The extent of adherence variation over time was rarely reported: Dean and Crittenden[48] reported a decrease in compliance rate from 100% to 87.7% in 12 months, and Min et al[36] observed a decline from 100% to 13.3% at 90 days. Other noteworthy observations were that noncompliers were slightly older,[43] that unemployed women had a higher compliance rate,[36] and that higher rates were seen for monthly rather than weekly adherence.[35] Snyder et al[11] reported that the proportion of missing PROM items was lower when PROMs were completed at home versus in the clinic.

Symptoms

Three randomized controlled trials (RCTs) evaluated the effect of PRO collection on symptom (severity) over time. All reported significantly decreasing symptom prevalence or symptom distress over time in the intervention arm versus controls.[32, 39, 49] Other studies[26, 41, 43, 45] also reported a reduction of symptoms postinter-

vention in comparison to baseline. More reporting of new or changing symptoms was observed when a PRO collection tool was used.[34] In addition, a (3-arm) RCT by Egbring et al[20] found that when PROs were administered through an application instead of through paper-based questionnaires, there was a higher frequency of symptom reporting. There was also a higher tendency to report overall symptoms when patients were not supervised by care providers.[20]

Health outcomes

Eleven studies evaluated the impact of PROM administration on patient well-being (Table 1). Reduction of psychosocial distress was reported,[30] as was reduction of symptom distress or burden,[42, 43] anxiety and depressive feelings,[26, 41] and pain.[27, 33] Furthermore, improved physical functioning and improved emotional and sexual well-being were found.[26, 33, 46, 48] A substantial effect on health-related quality of life over time between the PROM intervention and the control group was reported in 1 RCT (n = 776), in addition to the significantly higher overall and quality-adjusted survival in the intervention group.[24] Two other RCTs[26, 30] also reported improved health-related quality of life scores, albeit not significantly improved. A qualitative study,[46] in which focus groups and one-on-one interviews were conducted, found that for women postmastectomy, PROMs focusing on emotional well-being, education, communication, and process of care (e.g. scheduling appointments and transition of care) were of greatest importance.

Acceptability

Fifteen studies evaluated patients' acceptability of PRO collection (Table 1). Patients were generally positive, reporting that the interventions were helpful. Four studies used a tablet computer to collect PROMs, and over 90% of their study populations found tablets (very) easy to use.[22, 31, 33, 37] Similar results were reported for the use of applications[45, 50] and telephone-based interventions.[11, 41, 42] An online web tool was used in 4 studies,[19, 29, 31, 35] all of which reported an overall high acceptability. The majority of patients reported that the length and number of PROM questionnaires was acceptable,[19, 22, 37, 43] the questionnaires were helpful in discussing health issues with their care providers during visits,[11, 23, 29, 41, 47, 50] they were willing to answer additional questions,[23] and they liked being able to complete the questionnaires at home.[23, 29, 35, 47]

Satisfaction

Twelve studies reported the effect of PRO collection on patient satisfaction. A cohort study of 66 breast cancer patients that evaluated the use of a tablet computer for PROM collection found that 75% of patients were satisfied at baseline, and that percentage rose to 88% over time.[31] Of these patients, 84% to 94% were willing to recommend the PROM collection tool to other patients. In studies with a telephone-based PROM collection tool (n = 4), patients reported a high satisfaction rate as well.[11, 28, 41, 42] Comparable results were found for web-based collection tools, with patients reporting them as helpful and enjoyable.[19, 22] In 1 study, patients preferred the tablet computer over an online web tool.[23]

Impact of PROMs on Provider Outcomes

Compliance

Knoerl et al[19] examined providers' willingness to review PROMs: All providers (5 nurse practitioners and 1 physician assistant) created an account to a web-based platform (Carevive) for PROM collection and ultimately reviewed and finalized 81.3% of generated personalized care plans. However, only 20% of providers who completed all outcome assessments (n = 5) reported that they consistently reviewed care plans with their patients.[19]

Clinical decision making

Care providers' opinions were divided on whether PROM scores would influence clinical decision making. One study reported that none of the participating providers felt the PROM collection tool had led to different therapy decisions,[25] whereas another study reported that one third of providers indicated that their clinical decisions were influenced by symptom alerts.[32] Providers' email responses to symptom alerts were to maintain treatment strategies (46%), reassess the patient at the following clinic visit (33%), or alter the treatment (8%).[32]

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes.

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Abernethy et al.[31]**,** 2009	To determine feasibility and acceptability of tablet computers for administering standard assessment questionnaires and for collecting patient-reported symptom and QoL data	Prospective cohort study	66 breast cancer patients. Characteristics: - Mean age 54 y (6SD 12) - Female 100% - Caucasian 77% - Metastatic 61% - College degree or higher 86%	"eTablet", a tablet computer for PRO data collection - Paper & electronic (tablet computer) - Clinic waiting room - 4 times in 6 months	1. FACT-G 2. FACT-B 3. MDASI 4. FACIT-F 5. FACIT-Self-Efficacy Scale 6. PCM, an 86-item survey for common cancer- and treatment-related symptoms 7. Satisfaction & Acceptability survey
Albert et al.[25]**,** 2002	To implement individual QoL profile measurements in follow-up cancer care	Prospective cohort study	24 breast cancer patients Characteristics: - Median age 57 years (range 45-75) - Female 100%	Individualized graphic QoL Profiles based on PROMs - Electronic (software) - Clinic waiting room or doctor's office - At each follow-up visit	1. EORTC-QLQ-C30 2. EORTC-QLQ-BR23

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	- 4 clinicians reviewed a printed PRO summary report - 1 (research) nurse included PRO reports to patient files after marking alarming items	6	*Acceptability:* 60/64 (94%) patients found eTablets very easy to use, 98% found it easy to respond to survey questions on eTablets, 94% found the eTablet easy to navigate, 77% found the eTablet's weight very comfortable. *Satisfaction*: 48/64 (75%) patients were satisfied with the eTablet for PRO collection at baseline, and that percentage increased to 88% at the fourth visit; 84% of patients would recommend the PROM collection tool to others at the first visit, and that percentage increased to 94% by the fourth visit. *Other*: 42/57 (74%) patients at the first visit felt the PCM helped them recall experienced symptoms; a sentiment that remained in 29/40 (73%) patients at the fourth visit.	–	*Communication*: 16/50 (32%) of patients, at the first visit, felt encouraged to address symptoms with clinicians that they otherwise wouldn't have discussed; a sentiment that increased to 48% (16/33 patients) by the fourth visit
	14 clinicians during follow-up care received mailed QoL profiles with discharge and most recent QoL scores	–	–	*Compliance:* 9/14 (64%) clinicians actively participated by completing most (67% response rate) routine surveys about usefulness of QoL profiles *Clinical decision making:* None of the clinicians felt that the QoL profiles led to different therapy decisions *Acceptability:* - 16/16 (100%) found the QoL profiles understandable, 63% felt that they provided more information, 25% felt they helped notice patient issues - 94% felt that the profiles corresponded to their own patient assessments	*Communication:* 7/16 (44%) clinicians felt the QoL profiles had an effect on the patient-provider communication– namely that more specific questions were being asked

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Anderson et al.[32]**,** 2015	To determine the feasibility and efficacy of an automated telephone-based intervention on pain and symptom improvement in minority patients	RCT	60 low-income breast cancer patients Characteristics per group: *1) Intervention (n=31)*: - Mean age 49.6 years (± 9.9) - Female 100% - Latina 58% - Metastatic 35% - Years of education 10.6 years (± 4.1) *2) Control* (n=29): - Mean age 50.5 years (± 11) - Female 100% - Latina 59% - Metastatic 41% - Years of education 10 years (± 2.9)	*Intervention:* Automated IVR intervention for symptom reporting using dial-tone keypad for intensity rating - Telephone - Home - 2 times per week *Control:* Paper-based assessments at baseline and 2 follow-up points, without review by clinicians. Patients received usual symptom management, if indicated.	1. IVR-related pain and symptom list 2. MDASI 3. BQ-II 4. ECOG Performance Scale 5. PMI PROMs (#2 - #5) were administered during clinic visits at two time points (4-6 weeks and 8-10 weeks after enrollment).
Barbera et al.[40]**,** 2015	To determine the impact of screening with ESAS on unplanned ER visit rates of breast cancer patients	Retrospective cohort study	8359 breast cancer (stage I-III) patients receiving adjuvant chemotherapy within 6 months of diagnosis Characteristics per group: *1) With ESAS screening (n=2541):* - Mean age 53.22 years (± SD 10.44) *2) Without ESAS screening (n=5818):* - Mean age 53.87 years (± SD 10.57)	Screening with ESAS instrument - Electronic (web-based) and paper-based - Clinic waiting room - Each follow-up visit at cancer center	1. ESAS

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	Clinicians received IVR system-generated email alerts, when symptoms were moderate to severe (pain score ≥ 5 on a 0-10 scale)	2	*Adherence:* 71% of IVR assessments were completed successfully. *Symptoms:* - Proportion of patients with moderate to severe pain decreased from 86% to 43% (p=0.004). No significant decrease in control group (from 80% to 56%). - Proportion of patients in the intervention group with moderate to severe symptoms decreased significantly over follow-up (distress from 57% to 19%, p=0.008; sadness from 52% to 19%, p=0.04; drowsiness from 65% to 30%, p=0.04), respectively. Conversely, no significant changes were observed in the control group. - No significant difference in proportion of moderate to severe symptoms between intervention and control groups over time	*Clinical decision making*: - 33.3% of clinicians disclosed that their clinical decisions were influenced by symptom alerts. - Clinicians' email responses to symptom alerts were to maintain treatment course (46%), to assess the patient at the following clinic appointment (33%) or to prescribe a new symptom treatment (8%)	*Other*: - 100% of symptom alerts were detected; - 161/221 (73%) symptom alerts were acknowledged by clinicians through an email response to the research staff. - No significant difference between groups in proportion of patients receiving adequate analgesics (33% vs. 28%), as determined by the PMI
	Clinical team received summary reports on symptom scores to facilitate patient-provider communication about relevant symptoms	–	*Adherence:* Median number of ESAS assessments was 3 (IQR 1-5) for the 2541 patients that received at least 1 ESAS screening	–	*Hospital visits*: - ER visits were 43% lower in patients previously screened with ESAS compared with patients who were not; for each additional prior ESAS screening, there was a 17% decline in ER visits - Association of screening with ESAS on ER visits remained preventative even after adjusting for other types of visits

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Basch et al.[24], 2016	To determine the effect of routine web-based collection of PROMs on HRQoL and clinical outcomes in patients receiving adjuvant chemotherapy	RCT	143/776 (18.4%) breast cancer patients receiving outpatient chemotherapy Characteristics per group: *1) Intervention*: 89/441 (20%) breast cancer patients - 25% computer-experienced - 11% computer-inexperienced *2) Control*: 54/325 (17%) breast cancer patients - 19% computer-experienced. - 10% computer-inexperienced	*Intervention:* "STAR", a web-based interface for symptom reporting - Electronic (web-based; tablet computer) - Clinic and home (between-visit report) - Weekly email prompts to report symptoms *Control*: Usual care	1. CTCAE (adapted version) 2. EQ-5D

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	- Nurses received email alerts when symptoms progressed (by ≥ 2 points or absolute score ≥ 3 on a 0-5 scale) and responded directly - Clinicians received a printed report of symptoms tracked at each clinic visit	Median 3.7 (0.25-49)	*Adherence:* 73% of patients in the STAR arm completed a self-report at any given clinic visit *Health outcomes:* - HRQoL scores at 6 months improved in significantly more patients in the STAR group compared with control patients (34% vs. 18%, p<0.001) - EQ-5D subdomains mobility, self-care and anxiety/depression were significantly better in the STAR group compared to usual care (p=0.02, p=0.01 and p=0.01, respectively); significance was not reached for subdomains pain/discomfort and usual activities - Overall survival at 12 months was higher in the STAR arm compared to the control arm (75% vs. 69%, p=0.05). Mean 12-month quality-adjusted survival was also significantly higher in the STAR arm (8.7 vs. 8 months, p=0.004)	*Clinical decision making*: - 77% of alerts led to telephone counseling by nurses about symptom management, 12% of alerts led to start/change of supportive medication, 2% led to adjustment of chemotherapy dose	*Referrals:* 8% of alerts resulted in an ER/ hospital referral *Hospital visits:* Proportion of patients visiting the ER was significantly less in the STAR arm compared to usual care (34% vs. 41%, p=0.02) at 12 months A similar trend was observed for hospitalizations (45% vs. 49%, p=0.08) *Other:* - Patients in the STAR arm received significantly longer chemotherapy compared to the usual care group (8.2 vs. 6.3 months, p=0.002) - No significant difference in the amount of nursing calls was observed between the STAR arm and usual care (mean 12.8 vs. 12.9, p=0.93)

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Berry et al.[49], 2014	To determine the effect of a web-based PRO assessment and educational intervention on symptom distress during cancer treatment	RCT	206/752 (27.4%) breast cancer patients Characteristics per group: *1) Intervention*: 109/374 (30%) breast cancer patients *2) Control*: 97/378 (26%) breast cancer patients	*Intervention*: Combination of patient education, communication coaching and a PRO assessment - Electronic (web-based) - Home and clinic waiting room - At least 4 times (study time points) and voluntary between visits *Control*: Enhanced usual care with 4 PRO assessments with clinicians receiving a summary report	ESRA-C, an electronic patient-report application containing 3 PROMs: 1. SDS-15 2. PHQ-9 3. EORTC-QLQ-C30

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	Clinicians could be called immediately if symptom levels were severe between clinic visits and they received summary reports at each visit	–	*Adherence:* No significant difference in rates of outcome completion were observed between the intervention arm and usual care group (77.3% vs. 77.2%) *Symptoms:* - Symptom distress was significantly lower in the intervention arm vs. controls (mean change SDS-15 score 0.04 ± 5.8 vs. 1.27 ± 6.7) - A statistically significant reduction (average 1.93 score change in SDS-15, p=0.002) was observed in patients ≥ 50 years within the intervention arm compared to usual care. - No significant intervention effect was observed in patients < 50 years.	–	–

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Bock et al.[56], 2012	To determine the effect of an online health survey on symptom reporting, documentation and management by clinicians	Cross-sectional study	106 breast cancer patients Characteristics: - Mean age 56.9 years (32-85) - Female 100% - Caucasian 87% - Metastatic 0%	Online questionnaire on symptoms (frequency, severity and associated distress) and health behavior - Electronic (web-based with/without tablet computer) - Home or clinic waiting room	Unspecified PROM (symptoms and health history)

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	- 7 clinicians received patient-reported symptom reports attached to the medical record before clinic visit - 3 nurse practitioners	6	*Symptoms:* Number of patient-reported symptoms (mean 3.8, range 0-13) was significantly ($p<0.001$) higher than number of clinician-reported symptoms (mean 1.8, range 0-7)	–	*Other*: - More than half of symptoms mentioned by both patients and clinicians are addressed, regardless of number of symptoms - Considerable discordance between patient and clinician documentation of exercise behavior (100% vs. 28%) and alcohol use (100% vs. 70%)

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Børøsund et al.[39], 2014	To determine the effects of an online illness management system, IPPC service, and usual care on symptom distress, anxiety, depression and self-efficacy	RCT (3-arm)	167 breast cancer patients who underwent surgery or were receiving other treatments (radiation, chemotherapy, hormone therapy, or combinations) within 12 months post-surgery Characteristics per group: 1) *WebChoice intervention (n=64):* - Mean age 51 years (37-79) - Female 100% - College degree or higher 63% 2) *IPPC intervention: (n=45):* - Mean age 50 years (31-66) - Female 100% - College degree or higher 51% 3) *Usual care (n=58):* - Mean age 53 years (36-69) - Female 100% - College degree or higher 46%	*Intervention 1:* "WebChoice", an online illness management system *Intervention 2:* IPPC, a nurse-administered communication service - Electronic (web-based) - Home - Voluntary frequency *Control:* Usual care	1. Symptom list 2. MSAS 3. HADS 4. CBI

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	- 6 clinicians - 11 nurses - 3 social workers IPPC messages were answered primarily by nurses through secure e-messages	6	*Adherence:* 49/64 (77%) WebChoice users logged on at least once in 6 months; of those who logged on at least twice, median was 7 times (range 2-41). *Symptoms:* - Anxiety (mean difference -0.79, p=0.03) and depression (mean difference -0.79, p=0.03) were significantly lower in the WebChoice group vs. usual care group. - Symptom distress was significantly lower for patients in the WebChoice arm vs. the usual care arm (mean difference -0.16, p=0.001). No difference was observed in symptom distress between IPPC and usual care. - WebChoice participants had higher self-efficacy scores over time (mean difference 8.81, p=0.08) than the usual care group	Adherence: 33/153 (22%) IPPC messages were answered by clinicians	–

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Braeken et al.[30], 2013	To evaluate the short- and long-term effects of using a psychosocial screening instrument on psychological distress and HRQoL	Cluster RCT	284/568 (50%) breast cancer patients receiving radiotherapy Groups: 1) *Intervention*: 145/268 (54.1%) 2) *Control*: 139/300 (46.4%) (no patient characteristics described)	*Intervention:* Psychosocial screening instrument - Mailed - Home - 2 times (at 3 and 12 months follow-up) *Control*: Usual care	1. SIPP 2. HADS 3. GHQ-12 4. EORTC-QLQ-C30 5. VAS
Dean et al.[48], 2016	To determine the utility of the BREAST-Q as PROM in routine cancer care	Prospective cohort study	343 breast cancer patients Characteristics: - Median age 52 years (24-82)	Routine assessment with BREAST-Q instrument - Unspecified administration method - Clinic waiting room - Ideally 5 times	1. BREAST-Q

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	7 clinicians reviewed the questionnaire scores to get an overview of psychosocial issues and patient needs/ preference for psychosocial care	12	Results for entire study population (n=568): *Adherence:* > 85% of patients completed the study; no difference in loss-to-follow rates between intervention and control arms. *Health outcomes:* - No significant intervention effects on prevalence and extent of psychological distress were observed in short and long terms - No significant intervention effects on HRQoL were observed in short and long terms.	–	*Other*: Significant interactions between trial arm, 3-month follow-up, and referral rate were found: anxiety symptoms (β =2.16 and p = 0.03), emotional functioning (β = 15.16 and p = 0.02), appetite loss (β =15.67 and p = 0.04) and financial problems (β =11.39 and p=0.01). These interactions imply that early referral might affect short-term HRQoL.
	- Nurses who scored the questionnaire and transferred the scores to the database - Clinic clerks entered questionnaire data	24 (1-58)	*Adherence*: - 219/219 (100%) patients completed preoperative PROM prior to their first reconstructive procedure - 79/219 (36.1%) patients completed the PROM after undergoing all reconstructive procedures - 68/106 (64.2%) patients completed the minimum of three BREAST-Q assessments (preoperative, postoperative ≤ 3 months and post-completion) *Health outcomes*: At baseline (screening), patients with intact breasts had significantly (p<0.001) higher scores across all BREAST-Q domains than those with mastectomy defects. *Other*: Completion of BREAST-Q by patient took approximately 10 minutes	–	Other: - Data entry by clinic clerks of questionnaires took 3 minutes - Nurse needed 5 minutes per questionnaire to score and transfer scores to a database

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Decker et al.[41], 2009	To develop and test an AVR system for monitoring oral chemotherapy adherence and symptoms	Cohort study	24/30 (80%) breast cancer patients receiving oral chemotherapy Characteristics per group: *1) Adherence group (n=23, 77%):* - Breast cancer patients 74% - Age > 71 years: 30.4% / < 70 years: 69.6% - Female 91.3% - Caucasian 95.7% - College degree or higher 34.8% *2) Non-adherence group (n=7, 23%):* - Breast cancer patients 100% -Age < 70 years: 100% - Female 100% - Caucasian 71.4% - College degree or higher 42.3%	AVR system and self-help guide plus nursing intervention (for symptom management strategy) to monitor symptoms and oral chemotherapy adherence - Telephone (automated system) - Home - 8 times (weekly)	1. Symptom Experience Inventory 2. CESD-20 3. SF-12 4. Unspecified patient satisfaction questionnaire

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	Nurses would call patients when the AVR system indicated non-adherence and/ or symptom severity (≥ 4) for 3 consecutive weeks.	2.5	*Symptoms:* Difference in average sum of symptom severity before and after AVR intervention was a non-significant decrease of 4.35 (p=0.21). *Health outcomes:* - No significant differences in SF-12 items were observed between adherent and non-adherent group - Patients in the adherent group had a lower CESD-20 scores than patients in the non-adherent groups (8.67 vs. 11) *Acceptability:* 60% of patients felt the intervention was helpful, 30% felt that it was both burdensome and helpful, and 10% felt it was not helpful. *Satisfaction:* 17/17 (100%) patients that completed the questionnaire were either very satisfied or satisfied with the AVR system for monitoring symptoms. *Other:* 7/30 (23%) patients had confirmed non-adherence of chemotherapy	*Other:* Nurse interventions were especially indicated for fatigue and pain, the most commonly occurring symptoms with severity ≥4 for 3 consecutive weeks.	*Hospital visits:* 4/30 (13%) patients were admitted once to the hospital. 8/30 (27%) patients had primary care visits. *Other*: Out-of-pocket expenses for oral chemotherapy agents was not significantly different between the adherent and the non-adherent group.

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Egbring et al.[20], 2016	To determine the impact of a mobile app on patient-reported daily functional activity between supervised and unsupervised breast cancer patients	RCT (3-arm)	139 breast cancer patients receiving chemotherapy Characteristics per group: 1) *App use without physician review (unsupervised) (n=46):* - Mean age 50 years (± SD 10) - Female 100% - Metastatic 0% 2) *App use with physician review (supervised) (n=49):* - Mean age 53 years (± SD 12) - Female 100% - Metastatic 0% 3) *Control group (n=44):* - Mean age 56 years (± SD 15) - Female 100% - Metastatic 0%	*Intervention 1*: Mobile app for symptom reporting without clinician review (unsupervised) *Intervention 2*: Mobile app for symptom reporting with clinician review (supervised) - Electronic (mobile app and web-based) and paper-based - Home - Daily *Control*: Usual clinician support	1. ECOG Performance Status 2. CTCAE symptom list
Fu et al.[45], 2016	To describe the development and testing of TOLF health IT system, a web-based educational and behavioral intervention	Pilot cross-sectional study	30 breast cancer patients Characteristics: - Mean age 58.6 years (± SD 11.4) - Caucasian 73.3% - College degree or higher 86.6%	"TOLF" health IT system, an educational and behavioral self-care intervention including PROMs and symptom management - Electronic (mobile and web-based)	1. BCLE-SEI 2. Perceived Ease of Use and Usefulness Questionnaire 3. Post Study System Usability Questionnaire

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	Clinicians reviewed reported symptom data and patient charts of only supervised patients	1.5	*Symptoms:* - Both intervention groups reported more symptoms on the mobile app than on the paper-based questionnaire (supervised: 1033 vs. 656 symptoms; unsupervised 852 vs. 823 symptoms) More overall symptoms were reported by unsupervised patients vs. supervised patients (4808 vs. 4463 symptoms) *Health outcomes:* - Initially, all groups had a decline in functional activity scores from the first to the second visit; only supervised patients reported improvement of functional activity from the second to third visit. - Overall, supervised patients had a stable functional activity over from the first to the second (median 90.85 to median 84.76, p=0.72) *Satisfaction*: All supervised patients over time were satisfied with medical care	–	*Communication:* Fewer patients in the supervised group than the other groups had concentration problems during clinic visits; all supervised patients felt they were taken seriously
	–	3	*Acceptability:* 27/30 (90%) patients didn't have major usability problems with TOLF *Symptoms:* Patients reported significantly (p=0.022) less pain, tenderness, aching, soreness, and lymphedema at 12 weeks post-intervention compared to baseline	–	–

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Girgis et al.[26], 2009	To evaluate the effect of supportive care models on anxiety, depression, physical/ emotional functioning and unmet care needs	RCT	174/356 (49%) breast cancer patients Characteristics per group: *1) CATI with an oncologist/ general practitioner intervention:* - Breast cancer patients 33% - Mean age 58.3 years (37-75) - Female 72.3 % - College degree or higher 39.5% *2) CATI with telephone caseworker:* - Breast cancer patients 34% - Mean age 57.8 years (33-75) - Female 72.5 % - College degree or higher 31.7% *3) Controls* - Breast cancer patients 33% - Mean age 57.4 years (28-75) - Female 71.8% - College degree or higher 37.6%	"Supportive care model" including CATI with a telephone caseworker (intervention #1) or an oncologist/ general practitioner (intervention #2) - Telephone - Home - 3 times (baseline, at 3- and 6-month intervals) *Control*: Usual care	1. HADS 2. EORTC-QLQ-C30 3. SCNS-SF 4. Needs Assessment for Advanced Cancer Patient Questionnaire
Graf et al.[27], 2016	To identify variables that influence willingness of breast cancer patients to use electronic PROMs	Cross-sectional study	96 breast cancer patients Characteristics (n=96): - Median age 56.68 years (± SD 12.38) - Metastatic 68%	3-part survey concerning SES, HRQoL, and attitude towards electronic PROMs - Paper-based - Once	1. EORTC-QLQ-C30 2. EQ-5D-5L/ EQ-VAS 3. Validated partial questionnaires including modules of KPF-BK

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	- Clinicians received feedback reports from CATI's for discussion at following clinic visits - Nurses as telephone caseworkers	6	*Symptoms*: - No significant differences in prevalence of severe anxiety and depression between intervention and control groups. - No overall intervention effect over time on anxiety and depression was observed between groups; within the telephone caseworker group, there was a significant (p=0.01) decrease in elevated depression prevalence over time. *Health outcomes*: - Physical functioning was significantly (p=0.01) improved for patients in the telephone caseworker group. - No significant differences in QoL were observed between groups. - QoL scores improved over time within groups, but there was no significant (p=0.88) overall intervention effect on QoL over time. - No significant differences were observed between groups in emotional, cognitive, or social functioning. *Other*: - A trend towards decreased unmet supportive care needs was observed in the telephone caseworker group at 6 months (p=0.07). - No significant differences were seen in unmet needs between groups.	*Compliance:* Response to CATI feedback reports was significantly higher in the telephone caseworker group vs. oncologist/ general practitioner group (99.7% vs. 47.7%, p<0.0001)	*Referrals:* Referrals were significantly (p<0.0001) higher in the telephone caseworker group than the oncologist/ general practitioner group *Communication:* Patients who did CATI's with telephone caseworkers were significantly (p=0.0005) more inclined to strongly agree that study participation improved patient-doctor communication
	–	–	*Health outcomes:* Median EQ-VAS 64.67 (± SD 18.15), median EORTC-QLQ-C30 56.16 (± SD 23.56), overall QoL 58 (± SD 23.50). *Acceptability:* - 54% of patients welcomed electronic PROs - 47% of patients thought it would have a positive impact on care. *Other:* Scores in computer skills differed; 35% had advanced computer skills, 34% used tablets	–	–

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Hahn et al.[37], 2004	To describe the use and testing of a multimedia program for QoL assessment in cancer patients with low and high levels of literacy	Intervention study	50/126 (39.7%) breast cancer patients Characteristics (n=126): - Mean age 50.9 years (± SD 13.7) - Female 69.8% - Caucasian 29.4% - College degree or higher 60.3%	"Talking Touchscreen", a multimedia program - Electronic (tablet) - Once	1. FACT-G 2. SF-36 3. SGUQ
Holch et al.[21], 2017	To describe the development of eRAPID, a system for patient-report and online adverse event management during cancer treatment	Qualitative study	2/13 (15.4%) breast cancer patients Characteristics (n=13): - Mean age 53 years (35-69) - College degree or higher 84.6% 9 patient advocates from PRO group	"eRAPID", a system with an integrated web application interface and an online questionnaire builder (QTool) - Electronic (web-based)	PRAE, with responses being allocated to scores corresponding with the CTCAE severity grade and UKONS
Javid et al.[46], 2016	To assess breast cancer patients' and clinicians' perspective on measurement tools, timing and method of capturing PRO data	Qualitative study	15 post-mastectomy patients participated in focus groups or one-on-one interviews Characteristics: - Caucasian 85% - College degree or higher 69% 25 surgeons completed the prioritization questionnaire and participated in a web conference. A Stakeholder Advisory Panel (12 clinicians and 5 patients) reviewed input from patients and clinicians about breast surgery PRO collection.	Intervention: - - Other (focus groups) - Consenting patients received mailed prioritization questionnaires and surgeons received online surveys about PRO domains	1. Prioritization questionnaires about PRO domains across key time periods 2. Focus groups and one-on-one interviews, or web conference

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	–	–	*Adherence:* Individual item response was nearly 100%; *Acceptability:* - 95.2% reported that the touchscreen was easy or very easy to use, > 80% found the assessment not too long - 64.4% preferred using the touchscreen rather than having an interviewer ask the question (P=0.172) - 14.1% of patients would not be willing to do the survey each time when visiting the doctor	–	*Other:* Average PROM completion time differed significantly for low and high literacy patients (33 vs. 28 minutes, p=0.041)
	19 clinicians received email alerts when severe symptoms were reported, and could respond to alerts by viewing reports in the EHR.	–	*Adherence:* Patient compliance was monitored by tracking website visits and questionnaire completions. Adherence was encouraged by the system through weekly generated reminders (email/ text message) *Other:* Patients could securely report adverse events online from home.	–	*Other:* Automated tailored adverse event management advice for patients was generated by the system (from clinical algorithms)
	–	12	*Health outcomes*: - HRQoL (including concerns about treatment, treatment complications, treatment decision satisfaction, impact on family/friends) was consistently ranked the highest domain, whereas sexual function was the lowest. - Treatment decision-making and physical function were ranked highly preoperatively and short-term postoperatively. - Emotional wellbeing subdomains (fear of recurrence and impact on family and career) were prioritized highly in long-term postoperative period.	*Other*: - HRQoL domain (concerns about treatment, complications and decision satisfaction/regret) was prioritized highly in the preoperative and short-term postoperative period; impact on family and friends, and fear of recurrences was deemed most important in the long-term postoperative period. - In the late-term postoperative phase, providers also deemed emotional wellbeing (coping issues, distress feelings, personal relationships, support groups) a priority.	*Communication:* Communication was prioritized highly during the preoperative and short-term postoperative period. *Other*: Care process themes that came up during discussions were scheduling clinic appointments, team-based care, transition of care, nurse navigation, and continuity of care.

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Judson et al.[35], 2013	To determine long-term patient adherence rate with self-reporting of common symptomatic chemotherapy-related toxicities	Feasibility study within larger (3-arm) RCT	72/286 (25%) breast cancer patients receiving chemotherapy Characteristics (n=286): - Mean age 58 years (30-85) - Female 64% - Caucasian 88% - College degree or higher 85%	"STAR", a web-based interface for symptomatic toxicity reporting - Electronic (web-based) - Home	1. CTCAE 3. EQ-5D-5L 4. Performance status
Kelleher et al.[33], 2016	To determine how PROs of self-efficacy for selected symptoms were associated with symptom severity, and to examine differences in PROM administration	Observational study	65/178 (36.5%) breast cancer patients Characteristics (n=65): - Mean age 54.6 years - Caucasian 78% - Metastasic 60%	Multiple PROMs - Electronic (tablet) and paper-based - Clinic waiting room - Once	1. Arthritis Self-Efficacy Scale (modified version) 2. Chronic Pain Self-Efficacy Scale 3. MDASI ("pain" item) 4. FACT-G 5. Electronic PRO-Satisfaction
Kim et al.[44], 2016	To determine if a mobile mental-health tracker can potentially indicate depression, and to examine adherence on accuracy in depression screening	Cohort study	78 breast cancer patients Characteristics: - Mean age 44.35 years (± SD 7.01) - College degree or higher 52.6%	"Pit-a-Pat", a smartphone app for collecting PROs -Telephone (mobile app) - Daily (mental health) and bi-weekly (PHQ-9)	1. Mental health items: anxiety, mood and sleep satisfaction 2. PHQ-9

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	- Clinician received patients' printed STAR reports to review at each clinic visit - Nurse received triggered automated email alerts in case of high grade or worsening symptoms	8.5 (median)	*Adherence*: - In total, patients logged into STAR 8690 times, of which 71% from home (between visits) and 29% were log-ins at the clinic - Average monthly compliance was 83%; average weekly compliance was 62% - Self-reported reasons for non-adherence were as follows: 73% of patients said they forgot, were too busy, or did not feel like it, 11% experienced technical and illness-related barriers - Older age at baseline, Caucasian and higher education level were significantly associated with higher adherence	–	–
	–	–	*Health outcomes:* Self-efficacy for functioning, pain and other symptoms was associated with the reported outcomes of pain, FACT-G sub-scales and MDASI scales. *Acceptability*: Patients felt the tablet computer was easy to read, to use, to navigate, and comfortable to use *Satisfaction*: Mean satisfaction score was 19.9 (± SD 1.55), out of a possible score of 20	–	*Other:* No differences in patients' responses were found between methods of PROM administration (electronic vs. paper-based)
	–	11	*Adherence*: - Lower adherence group (n=58) reported 208/497 (42%) observations; the higher adherence group (n=20) reported 289/497 (58%) observations ($p<0.001$) - Self-report adherence was associated with an increase in the accuracy of depression screening, with all AUC's of the higher adherence group being statistically higher ($p<0.01$) than those of the lower adherence group *Symptoms*: Depression screening performance of mobile mental-health trackers was comparable to the traditional method, administration of a PHQ-9 test, in the clinical setting	–	*Other*: Screening accuracy with all 3 approaches (ratio, average, and frequency measurement of daily mental-health ratings) was statistically higher ($p<0.05$) for patients in the higher adherence group than for patients in the lower adherence group

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Knoerl et al.[19], 2017	To pilot test and determine the feasibility, acceptability, usability and satisfaction of the CPS among breast cancer patients and clinicians	Single arm, pre-/post-test prospective cohort study	25 breast cancer patients planned for or receiving neurotoxic chemotherapy Characteristics : - Female 100% - Caucasian 80% - Metastatic 36% - College degree 88%	"Carevive Planning System", a web-based platform for collecting patient-reported CIPN symptoms to generate a customized symptom care plan - Electronic (web-based platform) for symptoms - Clinic waiting room - 3 times (before clinic visit)	1. PRO-CTCAE 2. EORTC-QLQ-CIPN20 3. System Usability Scale 4. Adapted Acceptability E-scale

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	Providers (n=6) - 5 nurse practitioners - 1 physician assistant Characteristics: - Female 100% - College degree or higher 100% Providers reviewed generated personalized care plans based on patients' responses to PROMs and could edit the care plans to further tailor treatment options and referrals.	Unspecified; 3 clinic visits	*Adherence*: - 25/25 (100%) patients created a CPS account - All patients completed 74/75 (98.6%) questionnaires over 3 clinic visits *Acceptability*: - Easy to use 4.90 (± SD 0.29) (range 4-5) - Helpfulness of CPS 4.08 (± SD 0.93) (range 2-5) - Understandability of questions 4.75 (± SD 0.53) (range 3-5) - Acceptability of time required for PROM completion 4.58 (± SD 0.58) (range 3-5) - Overall acceptability ranged from 4.08 (± SD 0.93) to 4.90 (± SD 0.29) (range 1-5) (n=24) *Satisfaction*: - Enjoyment of CPS 4.19 (± SD 0.73) (range 3-5) - Overall satisfaction 4.56 (± SD 0.65) (range 3-5) *Other*: - Usability score 85.00 (± SD 11.54) (range 62-100)	*Adherence*: - 6/6 (100%) providers created a CPS account - 61/75 (81.3%) patient care plans were reviewed - 20% of providers reviewed care plans consistently with patients *Acceptability*: - Ability of CPS to identify appropriate issues of concern 3.20 (± SD 0.84) (range 2-4) - Ability to identify symptoms or areas of need 2.40 (± SD 1.14) (range 1-4) - Overall acceptability ranged from 1.60 (± SD 0.89) to 3.20 (± SD 0.84) (range 1-5) (n=5) *Other*: - Usability score 33.50 (± SD 17.19) (range 12.50-57.50) - 100% of CIPN care plan recommendations were accepted by providers - 35% of tasks associated with treatment recommendations were accepted at visit 1 and 53% of tasks were accepted at visit 3	*Communication*: - Helpfulness of CPS in guiding clinical conversations with patients 1.80 ± SD 1.10 (1-3) - Helpfulness in patient-provider communication improvement 1.60 ± SD 0.89 (1-3)

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Knoerl et al.[22], 2017	To determine if CPS can encourage patient activation in breast cancer patients for symptomatic polyneuropathy management	Single arm, pre-/post-test prospective cohort study	75 breast cancer patients planned for or receiving neurotoxic chemotherapy Characteristics: - Mean age 51.93 years (25-82) - Female 100% - Caucasian 88% - Metastatic 13.3%	"CPS", a web-based platform for collecting patient-reported CIPN symptoms to generate a customized care plan - Electronic (tablet computer) and paper-based (only PAM) - Clinic waiting room - 3 clinic visits NB. also a pencil/paper version of the PAM at the first and third visit	1. PRO-CTCAE (visit 1-3) 2. EORTC-QLQ-CIPN20 (visit 1-3) 3. PAM (visit 1 and 3) 4. System Usability Scale (visit 3) 5. Adapted Acceptability E-scale (visit 3)

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	Providers (n=6) - 5 nurse practitioners - 1 physician assistant Characteristics: - Female 100% - College degree or higher 100%	Unspecified; 3 clinic visits	*Symptoms*: - Mean QLQ-CIPN20 sensory and motor severity scores remained low across 3 study visits, ranging from 6.28 to 17.68 (scale 0–100) - Mean PRO-CTCAE numbness and tingling severity and interference scores also remained low, ranging from 0.32 to 1.07 (range 0–4) *Acceptability*: - Easy to use 4.49 (± SD 0.94) (range 1-5) *Other*: - 85% of patients possessed high (level III-IV) patient activation at baseline. PAM scores improved significantly (p=0.02) from baseline to final clinic visit from 67.15 (± SD 13.5) to 69.29 (SD ± 16.18) - Helpfulness of CPS 3.75 (± SD 1.10) (range 1-5) - Understandability of questions 4.27 (± SD 1.0) (range 1-5) - Acceptability of time required for PROM completion 4.03 (± SD 1.14) (range 1-5) - Overall acceptability ranged from 3.63 (± SD 1.18) to 4.49 (± SD 0.94) (range 1-5) *Satisfaction*: - Enjoyment of CPS 3.63 (± SD 1.18) (range1-5) - Overall satisfaction 4.11 (± SD 1.04) (range 1-5) *Other*: - Usability was highly rated with a mean score of 76.18 (± SD 21.93) (range 0-100)	*Acceptability*: Providers reported several barriers to reviewing the care plans with the patients: - lack of time to review - not finding CIPN management recommendations useful - difficulty logging into Carevive website due to password or software operational errors	*Other:* 60/67 (89.6%) patients received at least 1 care plan, while only 37/67 (55.2%) patients received all 3 care plans during the study period.

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Kuijpers et al.[47], 2015	To assess cancer survivors' and care providers' expectations of potential attributes of an interactive portal including PROs with feedback	Qualitative study	21/34 (61.8%) breast cancer survivors Characteristics: *Breast cancer survivors (n=21):* - Mean age 52.9 years (range 27-76) - Female 100% - College degree or higher 33% *Health professionals (n=31):* - Mean age 45.5 years (range 24-62) - Medical n=7 - Paramedical n=10 - Psychosocial n=14	Interactive portal including PROs with feedback, among others. - Other (focus groups) - Once	Focus group discussion on expectations' concerning an interactive portal
Leung et al.[38], 2016	To assess the feasibility of PROMIS CAT for fatigue and sleep-disturbance items	Cohort study	60/336 (17.9%) breast cancer patients Characteristics (n=336): - Mean age 57.44 years (± SD 15.71) - Female 54.8% - Caucasian 74.6% - Metastatic 30%	PROMIS CAT of cancer-related fatigue and other sleep-disturbance items - Electronic (tablet) - During clinic visits	1. PROMIS CAT 3 domain item banks 2. FACIT-F 3. ISI 4. Acceptability survey

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	–	–	*Other:* - Patients indicated that the information from completing PROMs could increase the knowledge about their health status - Patients had doubts about care providers' incentives to review PROs due to time constraints	*Other:* - Care providers expected an increase in knowledge of patients' health status - Care providers had their doubts about whom the responsibility for PROM review and feedback would fall onto	*Communication:* - Both patients and care providers expected that the patient-provider communication would improve through PROM feedback by providers (and thus better preparedness for appointment)
	–	–	*Acceptability:* - >98% indicated that symptom screening was not burdensome - 65% were willing to complete survey at every visit *Other:* - 67% thought PRO results were useful for clinicians	–	*Other*: - Average number of items completed of ISI was 5.36 (± SD 2.16) for sleep-disturbance, and 4.51 (± SD 1.59) for FACIT-F (score range 0-52) for fatigue - Overall time 15-20 minutes per patient to complete all measures

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Min et al.[36]**,** 2014	To determine the feasibility of a sleep-disturbance data collection app	Cohort study	38 Korean breast cancer patients receiving neoadjuvant chemotherapy 30 patients completed the study Characteristics (n=30): - Mean age 45 years (36-65) - Lymph node metastasis 77% - College degree or higher 47%	"Pit-a-Pat", a smartphone app for collecting PROs - Electronic (mobile app) - 30-minute interview at admission - Daily	30-minute interview (app use) plus: 1. EQ-5D-3L 2. BDI 3. Sleep disturbance symptoms 4. Acute symptomatic chemotherapy-related toxicity 5. Medication 'diary'

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	Clinicians could review the data in the EHR as the database was integrated in it.	3	*Adherence:* - Overall compliance rate 45% - Longitudinal compliance curve decreased from 100% (day 1) to 50% (day 34) to 13.3% (day 90) - Cumulative compliance rate decreased steadily from 50% (day 70) to 45% (day 90) - Unemployed women were associated with a higher rate of compliance (p=.03) - Rate of self-reporting in the jobless subgroup was significantly higher compared to employed patients (55% ± SD 25.7 vs. 30.7% ± 19.2, p=.006) *Acceptability:* Reasons that patients gave for not self-reporting were that the app did not work (38% of patients), forgetting to (29%), finding the app not useful (21%), feeling too sick to self-report (8%), not feeling like it (4%)	–	The rate self-reporting was higher in the subgroup with a 1-day lag time (self-report immediately after enrollment) than in patients that had a lag-time of ≥ 2 days (51.6% ± SD 24 vs. 29.6% ± 25.3, p=0.03)

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Mooney et al.[42], 2014	To determine whether timely emailed alert reports would encourage patient-provider communication, enhance unrelieved symptom treatment, and thus lead to better symptom outcomes	RCT	99/250 (39.6%) breast cancer patients receiving chemotherapy Characteristics per group: *1) Intervention (n=129):* - Breast cancer patients n=63 (48.8%) - Mean age 55.2 years (21-86) - Female 82.2% - Caucasian 89.6% *2) Control (n=121):* - Breast cancer patients n=36 (29.8%) - Mean age 55.8 years (19-81) - Female 69.4% - Caucasian 93.2%	*Intervention:* Automated IT-based chemotherapy-related symptom reporting system, with symptom alerts (if moderate-to-severe levels) being sent to care providers - Telephone - Home - Daily *Control:* Equivalent contact duration with automated IT-system without generated symptom alerts	10 symptoms (pain, fatigue, nausea/ vomiting, fever, insomnia, anxiety, depression, mouth sores, diarrhea, and constipation)

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	An oncology physician and nurse received automated alerts of unrelieved symptoms.	1.5	*Adherence*: - Overall daily call adherence was 65% of expected days - No difference in average days called between intervention and control groups (28.72 ± 15.62 days vs. 29.69 ± 16.78 days, p=0.66) *Symptoms*: - Most frequent moderate-severe symptoms in both groups were fatigue (89.2% of patients), trouble sleeping (74.9%) and pain (70.4%). - Severe levels were significantly (p<0.001) more common in the treatment group - Less talking about symptoms at patient-initiated contacts in intervention group (62 % vs. 73%, P=0.19) *Health outcomes*: - No significant difference between symptom severity or distress scores between arms (mean difference=0.06, p=0.58) *Acceptability*: - 94% quiet or very easy to use - 91% call length acceptable - 61% very much or quite helpful in keep track of their symptoms - 52% system helped them feel like participating in their care - 79% quiet or very confident the system notified their oncology providers of their symptoms - 25% agreed the system helped their doctor/nurse decrease their symptoms *Satisfaction*: - 77% quite or very satisfied with using the system	*Acceptability:* - 100% found the alert reports to be timely - 89% found the system easy to interpret - 82% found it useful - 0% reported technical difficulties *Satisfaction:* - 11/13 (85%) providers were somewhat to very satisfied with the system	*Hospital visits*: - No difference between intervention and control group in number of times an alert was generated (p=0.80) - No difference in frequency of patient- and/or provider-initiated unscheduled contacts (p=0.73) - No difference in patient- vs. provider-initiated unscheduled contacts (p=0.14) - More provider-initiated contacts in intervention group than the control group (18 vs. 10 contacts)

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Mooney et al.[43], 2017	To determine the efficacy of an automated symptom management system in reducing chemotherapy side-effects as vs. enhanced usual care	RCT	156/358 (43.6%) breast cancer patients, starting chemotherapy with at least 3 planned cycles Characteristics per group: *1) Intervention (n=180):* - Mean age 54.77 years (±12.17) - Female 75% - Caucasian 80% - College degree or higher 69% *2) Controls (n=178):* - Mean age 56.79 years (± 10.54) - Female 76% - Caucasian 86% - College degree or higher 69%	*Intervention:* "Symptom Care at Home" system, a combined intervention with daily symptom reporting, self-management coaching and nurse-practitioner follow-up - Telephone - Home - Daily *Control:* Enhanced usual care with daily symptom reporting without review by care providers	11 symptoms (fatigue, insomnia, nausea/vomiting, pain, numbness or tingling, depression, anxiety, distress over appearance, diarrhea, mouth sore, and trouble thinking/ concentrating)

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	Nurse practitioners responded to alerts indicating symptom severity exceeding present thresholds by calling intervention patients to provide guideline-based symptom care.	6 months or till chemo-therapy course completion, which ever came first	*Adherence*: - Daily call adherence was high and not significantly (p=0.80) different between intervention and control arms - 12% voluntary withdrawal was recorded but the difference between groups was not significant; non-compliant patients (SCH 25 patients; UC 27 patients) were slightly older (58.45 vs. 55.08 years, p=0.02) *Symptoms*: - Most prevalent symptoms with moderate-severe levels were fatigue (reported by 86% of patients), pain (80% of patients), sleep troubles (78% of patients) and nausea (60% of patients) - 10/11 reported symptoms were significantly lower for intervention patients (p<0.001 - 0.025) than control patients *Health outcomes*: - Post-baseline symptom burden reduction (treatment impact) for intervention patients was 3.59 severity points (p < 0.001), roughly 43% of the control group - Intervention patients had 3 times fewer (67% less) severe days (p<0.001) and 1.65 times fewer (39% less) moderate days (p=0.001) than control patients; intervention patients had 39% more mild days (p=0.016) and 25% more 'no symptom' days (p=0.006) than control patients *Other*: Intervention patients reported alerting symptoms only on 19% of calls, while control patients alerted on 37% of calls	–	–

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Ruland et al.[50], 2013	To report patients' use of WebChoice, a multi-component application for disease management in cancer patients	Cross-sectional study (sub-study of previously conducted RCT)	56/103 (54%) breast cancer patients, who had access to WebChoice (intervention) and had logged on at least 2 times Characteristics: - Mean age 51 years (± SD 7.1) - Metastatic 17.7% - College degree or higher 71%	"WebChoice", a multi-component interactive e-Health application including self-monitoring symptoms/ health problems, individualized information/advice for disease management, email communication with nurses, and a patient discussion forum. - Electronic (web-based)	1. 3 sections of WebChoice: self-management intervention, discussion forum, and personal email messages to a nurse 2. Diary (personal notes) 2. Usefulness questionnaire (scale 1-9 points)

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	Nurses monitored the accuracy of posts in the patient discussion forum, and also posted helpful advice.	6	*Adherence:* - System was used 4153 times (median=13.5 times, range 2-892 log-ons) - 216 personal messages to nurse (median=1 time) - Discussion forum most used of WebChoice with 374 posts by breast cancer patients, which was significantly (p=0.01) higher compared to 132 posts by prostate cancer patients - 291 vs. 209 (p=0.54) self-assessments were made by breast cancer and prostate cancer patients, respectively. - 93/103 (90%) of usefulness questionnaires were returned *Acceptability:* - Most useful section of WebChoice, considered by patients, was receiving an answer from the nurse (score 7.6, range 3-9) - Most common reasons for using each section were: "get help with a problem" (75% of breast cancer patients), "get information about a problem"(77%), compare experiences with other patients (81%) and "prepare communication with health personnel" (29%)	–	–

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Snyder et al.[28], 2010	To identify the topics that patients and clinicians report as being relevant in a PROM for use in clinical practice	Cross-sectional study	21/41 (51.2%) breast cancer patients Characteristics (n=41): - Mean age 63.8 years (± SD 12.56) - Female 51.12% - Caucasian 68.3% - College degree or higher 65.9%	Semi-structured interviews with patients and clinicians regarding prioritizing issues/ domains for discussion during appointments - Telephone - Home - Once	–

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	6/15 (40%) breast cancer clinicians (medical and radiation oncologists)	–	*Acceptability:* - 67% of breast cancer patients preferred questions about whether it is an issue they want to have addressed - Slight preference for a categorical assessment and reporting of results *Other:* - 2 domains discussed by >70% of breast cancer patients were pain and information needs - 7 domains discussed by <50% of all patients (n=41): cognitive function, social function, sexual function, nausea/vomiting, constipation, emotional function, and dyspnea - >70% of breast cancer patients reported physical function, role function, pain, fatigue and information needs, to be relevant for inclusion in a PROM	*Acceptability:* - 83% of breast cancer clinicians (n=6) preferred questions about how bothered the patient is *Other:* - 9 domains discussed by >70% of clinicians (n=15): physical function, nausea/vomiting, role function, diarrhea, information needs, pain, fatigue, constipation, and sexual function. - 3 domains discussed by <50% of clinicians: cognitive function, social function, and sleep problems. - >70% of clinicians reported 6 domains (information needs, cognitive function, emotional function, role function, sexual function, fatigue, fatigue) to be relevant for inclusion in a questionnaire	–

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continued*)

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Snyder et al.[11], 2013	To evaluate PatientViewpoint's use, usefulness, & acceptability to patients and clinicians	Prospective cohort study	34/47 (72%) breast cancer patients undergoing treatment (47/52 patients completed the study) Characteristics (n=37): - Mean age 57 years (range 28-80) - Caucasian 79% - Metastatic 52% - College degree or higher 73%	"Patient-Viewpoint", a web tool for administering PRO surveys at intervals and to generate those in graphical score reports - Electronic (web-based; integrated in EHR) - Home or clinic - Every 2 weeks (regardless of visit frequency)	1. 6 PROMIS short forms (physical function, pain interference, satisfaction with social roles, fatigue, anxiety, & depression) 2. EORTC-QLQ-BR23 3. 15-item feedback form on Patient-ViewPoint

(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
Clinicians could choose how to review PROM results: in paper report, on the website, in the EHR, or not at all.	–	*Adherence:* - 87% of questionnaires was completed offsite - 3/47 (60%) patients only completed PROMs in the clinic; 28/47 (60%) patients completed PROMs only at offset - 84% no missing items on PROMIS questionnaires when completed at home vs. 67% at the clinic *Symptoms:* Among breast cancer patients, most prevalent domains were systemic therapy including symptoms and/or side effects (53%), and sexual function (50%) *Acceptability:* - Patients were generally positive with >90% indicating "strong" usability - 46% reported clinicians used the reported information - 39% reported care quality improved *Satisfaction:* - Patients reported the webtool to be well organized and laid-out, and that it allowed opportunities to discuss issues that would otherwise have not been - Patients also mentioned that they questioned whether their care providers reviewed the PROM results	*Clinical decision making:* - Most likely to discuss systematic therapy (89%), pain interference (80%), fatigue (80%), and sexual function (6%) - Most common actions taken in response to identified issues were providing information and/or advice *Acceptability:* - PROM results of 24/47 (51%) patients were reviewed by clinicians in the HER, while in 17% of patients the results were reviewed on the paper report. Results of 15/47 (32%) patients were reviewed in PatientViewpoint or not at all - On 10/47 (21%) feedback forms, clinicians reported not using the PRO information. Clinicians reported the following reasons for questionnaire use: it provided additional information (51%), it confirmed knowledge of patients' problems (49%), it provided an overall assessment (43%), it identified issues for discussion (38%), it contributed to patient management (30%) - On 27/47 (58%) feedback forms, clinicians reported that the questionnaire assisted them in identifying areas of concern, and that it improved quality of care (54%) *Acceptability:* - Clinicians strongly preferred graphical PROM results in PatientViewPoint but preferred plain-text score reports in the HER - More explanation about PRO item content and score meaning was desired	*Other:* - PROMIS completion: median 6 minutes (2-12) - EORTC-QLQ-BR23 completion: median 3 minutes (1-11)

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Stover et al.[23], 2015	To develop a web-based PRO screening system through cognitive interviews with cancer patients, and to assess patients' and clinicians' acceptability and value of PRO review and feedback	2-phase qualitative study Phase 1: Cognitive interviews (9 breast cancer patients) Phase 2: Usability assessment in the clinic (10 breast cancer patients)	19/74 breast cancer patients (26%) receiving chemotherapy Characteristics (n=74): - Age ≥ 60 43% - Female 50% - Caucasian 66% - College degree or higher 38%	"Patient-Reported Symptom Monitoring" system - Electronic (web-based/ tablet) - Clinic waiting room	1. PRO-CTCAE 2. PROMIS- Global Health scale 3. 2 items written by the authors ("agenda setting" & "other unlisted symptoms")

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	A summarized report of most severe to least severe symptoms, was generated and given to both patients and clinicians prior to the appointment.	–	*Acceptability*: - 92% of patients found it helpful in discussing health issues - 82% wanted to review PRO results with clinicians during future appointments - 87% would recommend the system to other patients - 64% reported screening questions were helpful in discussing medical issues with provider that might have been missed otherwise - 80% chose to make an agenda for discussion during the visit - 92% were willing to answer additional questions - 82% were willing to do the survey at home *Satisfaction:* - High satisfaction with the web-based screening tool and summarized PRO report - Patients recommended the use of the tablet computer	*Acceptability:* - 83% of clinicians found summarized PRO report easy to interpret - 67% found it helpful for communicating with patients - 92% would recommend it to future patients - 92% found the PRO summarized report most useful for reviewing symptoms - 80% (4/5) felt that the PRO summary was (very) helpful in changing the treatment plan - 92% reported that discussing the summarized PRO report with their patient during the clinic appointment did not change the duration of the consultation	–

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Wheelock et al.[34]**,** 2015	To integrate and determine the use of SIS.NET, a multicomponent intervention, in the follow-up of breast cancer survivors	RCT	100 breast cancer patients (TNM stages I-III), who completed and recovered from acute treatment (surgery, chemo- and/or radiotherapy, or experimental therapy) Characteristics per group: *1) Intervention (n=59):* - Mean age 54.78 years (± SD 8.66) - Caucasian 69.5% *2) Control (n=41)*: - Mean age 53.32 years (± SD 10.79) - Caucasian 78.1%	*Intervention:* "SIS.NET", a combined intervention with routine online health surveys, generated summary reports with highlighted concerning symptoms, review by nurse practitioners, and tailored automated referrals, and additional patient requests for interim health surveys and appointments - Electronic (web-based) - Before clinic visit - 4 times (3-month intervals) *Control:* Standard care with similar online questionnaires with generated summary reports for clinician review during appointment	1. SF-36 2. PHQ-8 3. MSAS

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	Nurse practitioners received notifications from SIS.NET to review available completed questionnaires.	18	*Adherence:* No significant difference in proportion of questionnaires completed between intervention (average 50%) and control arms (average 62.5%) *Symptoms:* Significant difference in the number of changed or new symptoms between SIS.NET and control arm (mean 7.36 vs. mean 3.2, p=0.00445)	–	*Hospital visits:* No significant differences between intervention and control arms in number of appointments with clinicians (10.8 ± SD 8.2 vs. 9.6 ± SD 7.3, p=0.45) or number of breast cancer visits (4.2 ± SD 2.3 vs. 4.1 ± SD 1.8, p=0.78) *Other:* No significant difference in number of medical tests between SIS.NET and control arm (average 3.5 ± SD 2.2 vs. 3.8 ± SD 2.4, p=0.51)

Table 1 | PROMs Scored and Their Impact on Patient-, Clinician-, and Process- or System-Level Outcomes. (*continu*

Selected article	Study objective	Study design	Breast cancer study population	PROM intervention, method, setting and frequency of administration	Administered PROM
Wu et al.[29], 2016	To assess patients' and clinicians' perspectives on the usability of PatientViewpoint, a web tool to incorporate PROMs in clinical oncology practice	2-round quality improvement substudy	17/42 (40.5%) breast cancer patients Characteristics (n=42): - Mean age 65 years (32-83) - Metastatic 30% - Caucasian 81% - College degree or higher 69%	"PatientViewpoint", a web tool that allows clinicians to assign PROMs to patients, that collects PROM data, links it to the EHR, and displays it in graphical reports - Electronic (web-based) - Clinic or home or elsewhere - Every 2 weeks (regardless of visit frequency)	1.EORTC-QLQ- C30 2. SCNS-SF 3. PROMIS short forms

Note. *AVR* automated voice response; *BCLE-SEI*, Breast Cancer and Lymphedema—Symptom Experience Index; *BDI*, Beck Depression Inventory; *BQ-II*, Barriers Questionnaire-II; *BREAST-Q*, Breast-Questionnaire; *CAT*, Computerized Adaptive Testing; *CATI*, computer-assisted telephone interview; *CBI*, Cancer Behavioral Inventory; *CESD-20*, Center for Epidemiologic Studies Depression Scale-20; *EORTC-QLQ-CIPN20*, European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire for Chemotherapy-Induced Polyneuropathy Questionnaire-20 item; *CPS*, Carevive Planning System; *CTCAE*, Common Terminology Criteria for Adverse Events; *ECOG*, Eastern Cooperative Oncology Group; *EHR*, electronic health record; *EORTC-QLQ-BR23*, European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire of Breast, 23-items; *EORTC-QLQ-C30*, European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire of Cancer Patients, 30-item; *EQ-5D*, EuroQol 5-Dimension; *eRapid*, electronic patient self-Reporting of Adverse-events: Patient Information and aDvice; *EQ-5D-5L*, EuroQol 5-Dimension 5-Level; *EQ-VAS*, EuroQol-Visual Analogue Scale; *ER*, emergency room; *ESAS*, Edmonton Symptom Assessment Scale; *ESRA-C*, Electronic Self-Report Assessment-Cancer; *FACIT-F*, Functional Assessment of Chronic Illness Therapy—Fatigue Scale; *FACIT-Self-Efficacy Scale*, Functional Assessment of Chronic Illness Therapy—Self-Efficacy Scale; *FACT-B*, Functional Assessment of Cancer Therapy—Breast Cancer; *FACT-G*, Functional Assessment of Cancer Therapy—General; *GHQ-12*, Goldberg's General Health Questionnaire-12; *HADS*, Hospital Anxiety and Depression Scale; *HRQoL*, health-related quality of life; *IPPC*, Internet-based Patient-Provider Communication Service; *IQR*, interquartile range; *ISI*, Insomnia Severity Index; *IT*, information technology; *IVR*, interactive voice response; *KPF-BK*, Kölner Patientenfragebogen für Brustkrebs; *MDASI*, M. D. Anderson Symptom Inventory; *MSAS*, Memorial Symptom Assessment Scale; *PAM*, Patient Activation Measure; *PCM*, Patient Care Monitor; *PHQ-9*, Patient Health Questionnaire, 9-items; *PMI*, Pain Management Index; *PRAE*, Patient-Reported Adverse Events adaptation of the gold standard CTCAE; *PRO-CTCAE*, Patient-Reported Outcome version of the Common Terminology Criteria of Adverse Events; *PRO*, Patient-Reported Outcome; *PROM*, Patient-Reported Outcome Measure; *PROMIS*, Patient-Reported Outcomes Measurement Information System; *QoL*, quality of life; *RCT*, randomized controlled trial; *SCNS-SF*, Supportive Care Needs Survey; *SDS-15*, Symptom Distress Scale-15; *SES*, socioeconomic status; *SF-12*, Short Form-12; *SF-36*, Short Form-36; *SGUQ*, Standard Gamble Utility Questionnaire; *SIPP*, Dutch Screening Inventory of Psychosocial Problems; *SIS.NET*, System for Individualized Survivorship Care; *STAR*, Symptom Tracking and Reporting; *TNM*, Tumor Node Metastasis; *TOLF*, The-Optimal-Lymph-Flow health IT system; *UKONS*, United Kingdom Oncology Nursing Society; *VAS*, Visual Analogue Scale.

ued)

	(Medical) professional providing feedback to PRO	Intervention duration (months)	Patient outcomes	Provider outcomes	Care process/system outcomes
	Clinicians could review in the EHR an automatically generated graphical report of PROM data with poor or worsening outcomes highlighted.	–	*Acceptability:* Strengths of the webtool: - Notifications/ reminders to complete questionnaires - Possibility of filling out questionnaires at home - Discussing issues with clinicians that they otherwise would not have - Free-text comments for additional details - Possibility of tracking scores over time Recommendations: - Sufficient reminders about survey - Tailoring questions - Additional information on PRO results including consistent score meanings - Clinicians' responsibility to review scores, discuss abnormal results, and act on them	*Acceptability:* All clinicians reported that it could be helpful Concerns: - Limited time to review PRO results if they were filled in just before the appointment - Rather face-to-face interaction than looking at graphical reports in EHR Wanted: - Email reminders - Graphical presentation is preferred above tables - Clearer and consistent score meanings for interpretation - Full integration into her - No additional login to access system	–

Acceptability

Five studies evaluated whether PROMs were helpful to providers in identifying and controlling patient symptoms and in other areas of need.[11, 22, 23, 25, 42] A prospective cohort study[11] reported that 58% of providers were most likely to agree that a PROM intervention helped them to identify areas of concern. Stover et al,[23] who implemented a web-based PRO screening system with 21 items assessing symptoms and functional status, reported that providers found PRO summary reports most useful for reviewing symptoms (92%) and (very) helpful in changing the treatment plan (80%). Most providers (92%) also reported that discussing PRO results with patients during clinical visits did not lengthen the consultation time. Albert et al[25] evaluated an electronic PRO collection tool (QoL Profiles) and reported that providers (n = 14) found the system understandable (100%), felt it provided more information (63%), helped them notice patients' issues (25%), and found that the system corresponded to their own patient assessment (94%). Knoerl et al[19] also observed that care providers reported the PROM collection to be helpful in identifying appropriate areas of concern or need.

Impact of PROMs on Care Process Outcomes

Referrals

There was limited evidence concerning the effects of monitoring or tracking PROMs on referral rates. Girgis et al[26] found a significant ($P < .0001$) difference in the number of referral recommendations as a response to PROM results between nurses and clinicians, with more referrals being recommended by nurses for unmet psychological, daily living, health service or information, and physical needs. Basch et al[24] demonstrated that 8% of severe symptom email alerts resulted in referrals, whereas the majority (77%) of email alerts prompted telephone counseling by nurses for symptom management.

Hospital visits

Only 2 studies looked specifically at the effect of PRO collection and screening on the rate of emergency room (ER) visits.[24, 40] Basch et al[24] reported a significant 7% decrease ($P = .02$) in ER visits between patients whose symptoms were electronically tracked and patients who received usual care during active chemotherapy. Barbera et al[40] reported comparable results, with the number of ER visits being 43% lower in breast cancer patients who received routine PRO assessments prior to clinic visits during active chemotherapy as opposed to patients who did not. Effects of PROM interventions on hospitalizations were not significant: Basch et al[24] found a 4% decrease in hospitalizations ($P = .08$) in patients whose PROMs

were monitored compared with patients with usual care; Wheelock et al[34] did not find a difference between patients whose PROMs were electronically captured during breast cancer follow-up care and patients who did not receive the PROM intervention.

Communication

Regardless of study design, patients and providers were generally positive about the impact of systematic PROM collection on patient–provider communication.[20, 25, 26, 31] Patients felt more encouraged after PRO assessment to discuss symptoms and issues with their providers that they otherwise would not have discussed[31] and also felt more focused and taken seriously during the consultation.[20] Although evidence was scarce, providers reported that reviewing PROMs encouraged clinical interactions between them and their patients[19] and that this review also led to more specific questioning during the clinic visit.[25]

Facilitators and Barriers

Facilitators

Technical facilitators of PROM implementation in routine breast cancer care were the most frequently described (Figure 2). One article pointed out that technological methods of PROM collection (e.g. web-based, smartphone applications, tablet computer) might be more acceptable.[32] Providers seemed to prefer graphic presentations of PRO results over table reports, but they also valued additional free-text comments by patients. Patient feedback included the recommendation of tailoring questions (through an interface) for individual patients and providing more information on PRO score results, which might enhance PRO collection and interpretation.[11, 29] Email reminders to review PRO results could possibly increase care providers' awareness of patient symptoms and problems and thus encourage symptom management.[32] Some patients expressed a preference to be notified when providers reviewed PRO reports.[29] One study found that the presence of a "change champion" was essential to facilitating PROM implementation by not only encouraging patients' acceptance of the collection method (e.g. tablet computers), but also guiding care providers through PRO reports and ensuring that appropriate actions were taken in response to significant patient issues.[31]

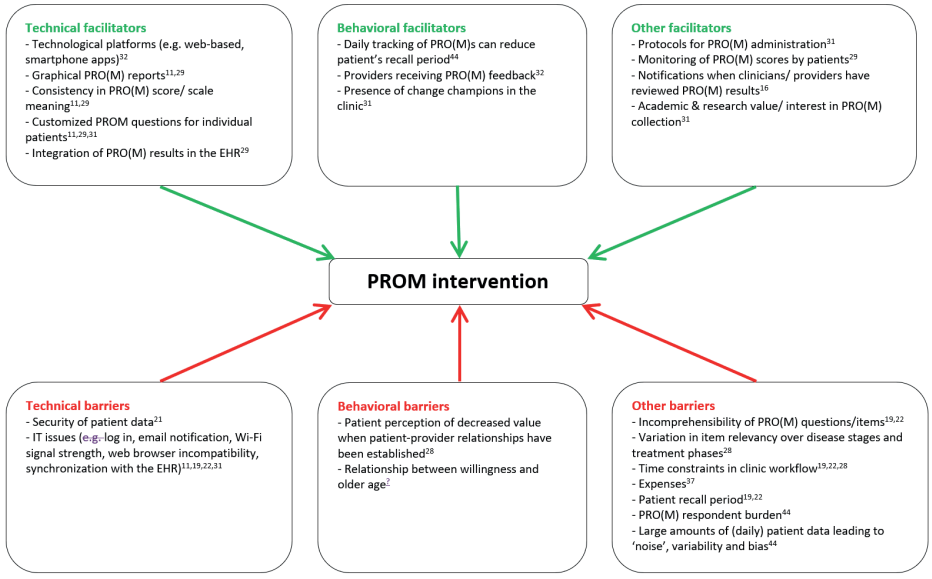


Figure 2 | Description of Facilitators and Barriers.
Note. EHR indicates electronic health record; IT, information technology; PRO(M), patient-reported outcome (measures).

Barriers

The clarity of PROM questions, including stating the (symptom) recall period, can be a hurdle for patient acceptability of the PROM intervention.[19, 22] Providers also indicated that the lack of integration and synchronization of the PRO data system with the electronic health record was a significant barrier in routine PRO review, because it required logging into additional systems (with additional passwords) and compatible Internet browsers.[11, 19, 22, 29] Some articles reported that patients found it confusing if the meaning of PRO summary scores was inconsistent, with higher scores alternatively indicating better or worse outcomes.[11, 29] Conflicting providers' opinions were reported regarding the timing of PROM review: Some providers indicated that most symptom alerts were received outside the hospital, thus impeding swift patient contact, whereas others reported that reviewing PROMs right before clinic appointments might disrupt the clinical practice workflow owing to time constraints.[19, 22, 29, 32] Another feasibility aspect was the burden to patients and staff: Not only can multiple PROMs become burdensome for patients over time, but reviewing and interpreting enormous amounts of patient data can also take its toll on providers.[28, 31, 44]

DISCUSSION

This review studied implementation methods, impact, and facilitators and barriers of PROM collection in breast cancer clinical practice. The PROM interventions were generally developed to improve symptom management, to identify psychosocial problems, to facilitate patient-centered care and treatment-specific monitoring during treatment phases, and to improve patient-provider communication. Electronic PRO collection systems were most often used, which can allow patients to efficiently complete standardized assessments, can increase usability with a lower response burden, can lead to fewer missing data (when compared with paper-based PROMs), and can lead to higher patient satisfaction.[31, 51, 52] Electronic PROM collection can also provide a certain flexibility in assessment location (clinic vs home) and frequency and can bridge treatment-specific and long-term PRO follow-up.[6]

An essential point is that some selected studies described the effects of a combined intervention, which included not only a PROM collection component but also additional actions, such as enhanced care with patient education and coaching, proactive review of PROMs, and follow-up by providers. This makes it difficult, from a practical standpoint, to determine the extent to which the observed impact of these multicomponent interventions can be attributed specifically to PROM collection and review. Then again, it might be unnecessary to separate the components, considering that complex interventions can lead to synergistic results and thus to an accumulated impact. It might be more important to determine the degree to which the individual parts of multicomponent strategies are implemented properly and according to the original proposed theory of the intervention.[53]

Apart from differences in PROM collection methods across selected studies, there was also notable variety in the frequency ("intensity") of the assessments: PRO assessments ranged from daily to once (for screening purposes). PRO collection tools were used to monitor either specific treatment phases (e.g. adverse effects during chemotherapy radiotherapy) or entire breast cancer care trajectories (from diagnosis to the end of follow-up). Successful implementation of routine PRO assessment should integrate both options and should combine them with the appropriate PROM to ensure optimal patient engagement and management of care.

The heterogeneous study populations should be taken into account; more than half of the selected studies did not have a study population consisting entirely of breast cancer patients. In the studies included in our review, the percentage of

breast cancer patients varied from 17.9% to 80%, with breast cancer patients being the majority in only 9 studies. Ruland et al.[50] described the differences between breast and prostate cancer patients in their use of WebChoice, an interactive e-Health application designed to support cancer patients in illness management. Breast cancer patients used the application twice as often and asked significantly more often for help than did prostate cancer patients. The gender-specific nature of different cancer types makes it difficult to identify attributes that could explain usage discrepancies. Nevertheless, the findings are consistent with previous research, which indicates that women seek more health-related information online than men do. In addition, the breast cancer patient population in the selected studies was overall younger than other cancer populations, and it is well known that younger people seek information online more often than older people do.[54, 55]

Another striking observation is the limited diversity in study populations across the selected articles. Most of the included study participants were overwhelmingly Caucasian.[11,19,22,23,28,29,31,33,34,35,38,41,42,43,46,56] There were only 2 studies in which PROM interventions were targeted at predominantly minority populations.[32,37] Limited diversity also applied to education levels of study participants across most included studies, with the majority of patients having a college degree or higher. As most of the interventions across the included studies required patients to be able to speak English, it became obvious that routine PROM collection was primarily executed in a select patient population. Another important point is that most patients across the included articles were recruited in clinic waiting rooms by being asked if they were willing to participate. This could have introduced some selection bias if willingness itself has associations with better adherence to other care components. Naturally, this selectiveness in patient characteristics can affect the outcome results of PROM interventions. By striving to avoid "preselection" of patients to participate in PROM intervention studies, generalizability issues can also be (partially) circumvented in future research.

Numerous selected studies in this systematic review mentioned facilitators and barriers of the PROM interventions. It must be noted that nearly all of the selected studies did not identify these facilitators and barriers after a thorough analysis because that was not the primary aim of the selected articles. The facilitators and barriers were mostly identified by observing the effects of PROM implementation in practice. In a systematic review about facilitators and barriers of PROM implementation in palliative care, Antunes et al.[57] reported that substantial leadership is necessary to boost motivation in patients and care providers and that educating them on the use, interpretation, and advantages of routine PROM collection could

increase adherence. Further research involving facilitator and barrier analysis should be conducted to provide more guidance during the design of PROM interventions.

Among the 2311 initially screened articles, 8 reviews were excluded based on irrelevant aims (see Methods). Their references, however, were also screened by title and abstract to ensure thoroughness of the literature search. Fourteen references[12,15,58,59,60,61,62,63,64,65,66,67,68,69] were identified as potentially relevant but were not among the initial 2311 articles. The apparent reason was that these articles were not indexed with breast cancer-related keywords or medical subject headings (e.g. "breast cancer," "breast tumor," "breast neoplasm"), probably because their study populations contained only a minimal proportion of breast cancer patients. So, it must be noted, for transparency reasons, that there could be some relevant articles that were not included because of improper/ incomplete indexing. These 14 articles did not provide any novel insights.

CONCLUSION

PROMs can be collected in a variety of methods, locations, and frequencies. Although interpreting the impact of PROM collection is challenging due to considerations of synergistic (multicomponent) interventions and generalizability issues, this review found that systematic PROM collection in routine breast cancer care has a promising impact on patients, providers, and care processes/ systems.

RECOMMENDATIONS

A major challenge in the implementation of value-based healthcare is standardizing PROMs that are meaningful to patients across different cultural and geographical settings.[70] Although there seems to be skepticism around the feasibility and effectiveness of standardized PROM collection as a part of breast cancer care, a number of PROMs (the BREAST-Q; the 30-item European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire of Cancer Patients; the 23-item European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire of Breast; the EuroQol 5-Dimension, 5-Level; and the Functional Assessment of Cancer Therapy-Endocrine Subscale) have been validated within the breast cancer population.[71] A breast cancer outcome set (including some of the aforementioned PROMs) was developed by the International Consortium for Health

Outcomes Measurement, which recommended a timeline for PRO assessments to evaluate certain events (baseline, after neoadjuvant chemotherapy, surgery, and follow-up) in breast cancer care.[71] If routine PROM collection is to be implemented in clinical practice, we recommend a standardized PROM set because it is one of the requirements for benchmarking treatments and healthcare providers. It could also ultimately allow comparisons of breast cancer outcomes across countries. By doing so, a major step will have been taken toward value-based healthcare.

ACKNOWLEDGMENTS

We acknowledge Gerdien B. de Jonge (Biomedical Information Specialist, Medical Library, Erasmus MC) for her support in the literature search. Authors AO, LMV, and JAH were supported by a grant from the Federation of Dutch University Medical Centers.

REFERENCES

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin*. 2011;61:69–90.
2. Porter ME, Larsson S, Lee TH. Standardizing patient outcomes measurement. *N Engl J Med*. 2016;374:504–506.
3. Bredart A, Marrel A, Abetz-Webb L, Lasch K, Acquadro C. Interviewing to develop patient-reported outcome (PRO) measures for clinical research: eliciting patients' experience. H*ealth Qual Life Outcomes*. 2014;12:15.
4. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. *J Eval Clin Pract*. 2006;12:559–568.
5. Snyder CF, Aaronson NK. Use of patient-reported outcomes in clinical practice. Lancet. 2009;374:369–370.
6. Jensen RE, Snyder CF, Abernethy AP, et al. Review of electronic patient-reported outcomes systems used in cancer clinical care. *J Oncol Pract*. 2014;10:e215–e222.
7. Trotti A, Colevas AD, Setser A, Basch E. Patient-reported outcomes and the evolution of adverse event reporting in oncology. *J Clin Oncol*. 2007;25:5121–5127.
8. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. *BMC Health Serv Res.* 2013;13:211.
9. Basch E, Artz D, Iasonos A, et al. Evaluation of an online platform for cancer patient self-reporting of chemotherapy toxicities. *J Am Med Inform Assoc*. 2007;14:264–268.
10. Basch E, Artz D, Dulko D, et al. Patient online self-reporting of toxicity symptoms during chemotherapy. *J Clin Oncol*. 2005;23:3552–3561.
11. Snyder CF, Blackford AL, Wolff AC, et al. Feasibility and value of PatientViewpoint: a web system for patient-reported outcomes assessment in clinical practice. *Psychooncology*. 2013;22:895–901.
12. Boyes A, Newell S, Girgis A, McElduff P, Sanson-Fisher R. Does routine assessment and real-time feedback improve cancer patients' psychosocial well-being? *Eur J Cancer Care (Engl)*. 2006;15:163–171.
13. Rosenbloom SK, Victorson DE, Hahn EA, Peterman AH, Cella D. Assessment is not enough: a randomized controlled trial of the effects of HRQL assessment on quality of life and satisfaction in oncology clinical practice. *Psychooncology*. 2007;16:1069–1079.
14. Velikova G, Booth L, Smith AB, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. *J Clin Oncol*. 2004;22:714–724.
15. Hilarius DL, Kloeg PH, Gundy CM, Aaronson NK. Use of health-related quality-of-life assessments in daily clinical oncology nursing practice: a community hospital-based intervention study. *Cancer*. 2008;113:628–637.
16. Basch E, Deal AM, Dueck AC, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. *JAMA*. 2017;318:197–198.
17. Howell D, Molloy S, Wilkinson K, et al. Patient-reported outcomes in routine cancer clinical practice: a scoping review of use, impact on health outcomes, and implementation factors. *Ann Oncol*. 2015;26:1846–1858.

18. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med*. 2009;151:W65–W94.
19. Knoerl R, Dudley WN, Smith G, Bridges C, Kanzawa-Lee G, Lavoie Smith EM. Pilot testing a web-based system for the assessment and management of chemotherapy-induced peripheral neuropathy. *Comput Inform Nurs*. 2017;35:201–211.
20. Egbring M, Far E, Roos M, et al. A mobile app to stabilize daily functional activity of breast cancer patients in collaboration with the physician: a randomized controlled clinical trial. *J Med Internet Res*. 2016;18:e238.
21. Holch P, Warrington L, Bamforth LCA, et al. Development of an integrated electronic platform for patient self-report and management of adverse events during cancer treatment. *Ann Oncol.* 2017;28:2305–2311.
22. Knoerl R, Lee D, Yang J, et al. Examining the impact of a web-based intervention to promote patient activation in chemotherapy-induced peripheral neuropathy assessment and management. *J Cancer Educ*. 2018;33:1027–1035.
23. Stover A, Irwin DE, Chen RC, et al. Integrating patient-reported outcome measures into routine cancer care: cancer patients' and clinicians' perceptions of acceptability and value. *EGEMS (Wash DC).* 2015;3:1169.
24. Basch E, Deal AM, Kris MG, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol. 2016;34:557–565.
25. Albert US, Koller M, Lorenz W, et al. Quality of life profile: from measurement to clinical application. *Breast*. 2002;11:324–334.
26. Girgis A, Breen S, Stacey F, Lecathelinais C. Impact of two supportive care interventions on anxiety, depression, quality of life, and unmet needs in patients with nonlocalized breast and colorectal cancers. *J Clin Oncol*. 2009;27: 6180–6190.
27. Graf J, Simoes E, Wisslicen K, et al. Willingness of patients with breast cancer in the adjuvant and metastatic setting to use electronic surveys (ePRO) depends on sociodemographic factors, health-related quality of life, disease status and computer skills. *Geburtshilfe Frauenheilkd*. 2016;76:535–541.
28. Snyder CF, Jensen RE, Geller G, Carducci MA, Wu AW. Relevant content for a patient-reported outcomes questionnaire for use in oncology clinical practice: putting doctors and patients on the same page. *Qual Life Res*. 2010;19:1045–1055.
29. Wu AW, White SM, Blackford AL, et al. Improving an electronic system for measuring PROs in routine oncology practice. *J Cancer Surviv*. 2016;10:573–582.
30. Braeken APBM, Kempen GIJM, Eekers DBP, et al. Psychosocial screening effects on health-related outcomes in patients receiving radiotherapy. A cluster randomised controlled trial. *Psychooncology*. 2013;22:2736–2746.
31. Abernethy AP, Herndon IJE, Wheeler JL, et al. Feasibility and acceptability to patients of a longitudinal system for evaluating cancer-related symptoms and quality of life: pilot study of an e/Tablet data-collection system in academic oncology. *J Pain Symptom Manage*. 2009;37:1027–1038.
32. Anderson KO, Palos GR, Mendoza TR, et al. Automated pain intervention for underserved minority women with breast cancer. *Cancer*. 2015;121:1882–1890.
33. Kelleher SA, Somers TJ, Locklear T, Crosswell AD, Abernethy AP. Using patient reported outcomes in oncology clinical practice. *Scand J Pain*. 2016;13:6–11.

34. Wheelock AE, Bock MA, Martin EL, et al. SIS.NET: a randomized controlled trial evaluating a web-based system for symptom management after treatment of breast cancer. *Cancer*. 2015;121:893–899.
35. Judson TJ, Bennett AV, Rogak LJ, et al. Feasibility of long-term patient selfreporting of toxicities from home via the Internet during routine chemotherapy. J Clin Oncol. 2013;31:2580–2585.
36. Min YH, Lee JW, Shin YW, et al. Daily collection of self-reporting sleep disturbance data via a smartphone app in breast cancer patients receiving chemotherapy: a feasibility study. *J Med Internet Res*. 2014;16:e135.
37. Hahn EA, Cella D, Dobrez D, et al. The talking touchscreen: a new approach to outcomes assessment in low literacy. *Psychooncology*. 2004;13:86–95.
38. Leung YW, Brown C, Cosio AP, et al. Feasibility and diagnostic accuracy of the Patient-Reported Outcomes Measurement Information System (PROMIS) item banks for routine surveillance of sleep and fatigue problems in ambulatory cancer care. *Cancer*. 2016;122:2906–2917.
39. Børøsund E, Cvancarova M, Moore SM, Ekstedt M, Ruland CMl. Comparing effects in regular practice of e-communication and Web-based self-management support among breast cancer patients: preliminary results from a randomized controlled trial. *J Med Internet Res*. 2014;16:e295.
40. Barbera L, Sutradhar R, Howell D, et al. Does routine symptom screening with ESAS decrease ED visits in breast cancer patients undergoing adjuvant chemotherapy? *Support Care Cancer*. 2015;23:3025–3032.
41. Decker V, Spoelstra S, Miezo E, et al. A pilot study of an automated voice response system and nursing intervention to monitor adherence to oral chemotherapy agents. *Cancer Nurs*. 2009;32:E20–E29.
42. Mooney KH, Beck SL, Friedman RH, Farzanfar R, Wong B. Automated monitoring of symptoms during ambulatory chemotherapy and oncology providers' use of the information: a randomized controlled clinical trial. *Support Care Cancer*. 2014;22:2343–2350.
43. Mooney KH, Beck SL, Wong B, et al. Automated home monitoring and management of patient-reported symptoms during chemotherapy: results of the symptom care at home RCT. *Cancer Med*. 2017;6:537–546.
44. Kim J, Lim S, Min YH, et al. Depression screening using daily mental-health ratings from a smartphone application for breast cancer patients. *J Med Internet Res*. 2016;18:e216.
45. Fu MR, Axelrod D, Guth AA, et al. mHealth self-care interventions: managing symptoms following breast cancer treatment. Mhealth. 2016;2. https://doi.org/10.21037/mhealth.2016.07.03 [Epub ahead of print].
46. Javid SH, Lawrence SO, Lavallee DC. Prioritizing patient-reported outcomes in breast cancer surgery quality improvement. *Breast J*. 2017;23:127–137.
47. Kuijpers W, Groen WG, Loos R, et al. An interactive portal to empower cancer survivors: a qualitative study on user expectations. *Support Care Cancer*. 2015;23:2535–2542.
48. Dean NR, Crittenden T. A five year experience of measuring clinical effectiveness in a breast reconstruction service using the BREAST-Q patient reported outcomes measure: a cohort study. *J Plast Reconstr Aesthetic Surg*. 2016;69:1469–1477.
49. Berry DL, Hong F, Halpenny B, et al. Electronic self-report assessment for cancer and self-care support: results of a multicenter randomized trial. *J Clin Oncol*. 2014;32:199–205.

50. Ruland CM, Maffei RM, Børøsund E, Krahn A, Andersen T, Grimsbø GH. Evaluation of different features of an eHealth application for personalized illness management support: cancer patients' use and appraisal of usefulness. *Int J Med Inform*. 2013;82:593–603.
51. Kongsved SM, Basnov M, Holm-Christensen K, Hjollund NH. Response rate and completeness of questionnaires: a randomized study of Internet versus paper-and-pencil versions. *J Med Internet Res*. 2007;9:e25.
52. Bliven BD, Kaufman SE, Spertus JA. Electronic collection of health-related quality of life data: validity, time benefits, and patient preference. *Qual Life Res*. 2001;10:15–22.
53. Alexander JA, Hearld LR. What canwe learn from quality improvement research? A critical review of research methods. *Med Care Res Rev*. 2009;66:235–271.
54. Owen JE, Goldstein MS, Lee JH, Breen N, Rowland JH. Use of health-related and cancer-specific support groups among adult cancer survivors. *Ca*ncer. 2007;109:2580–2589.
55. Kummervold PE, Chronaki CE, Lausen B, et al. eHealth trends in Europe 2005-2007: a population-based survey. *J Med Internet Res*. 2008;10:e42.
56. Bock M, Moore D, Hwang J, et al. The impact of an electronic health questionnaire on symptom management and behavior reporting for breast cancer survivors. *Breast Cancer Res Treat*. 2012;134:1327–1335.
57. Antunes B, Harding R, Higginson IJ, EUROIMPACT. Implementing patient-reported outcome measures in palliative care clinical practice: a systematic review of facilitators and barriers. *Palliat Med*. 2014;28:158–175.
58. Berry DL, Blumenstein BA, Halpenny B, et al. Enhancing patient-provider communication with the electronic self-report assessment for cancer: a randomized trial. *J Clin Oncol*. 2011;29:1029–1035.
59. Carlson LE, Groff SL, Maciejewski O, Bultz BD. Screening for distress in lung and breast cancer outpatients: a randomized controlled trial. *J Clin Oncol*. 2010;28:4884–4891.
60. Detmar SB, Muller MJ, Schornagel JH, Wever LD, Aaronson NK. Health-related quality-of-life assessments and patient-physician communication: a randomized controlled trial. *JAMA*. 2002;288:3027–3034.
61. Ganz PA, Greendale GA, Petersen L, Zibecchi L, Kahn B, Belin TR. Managing menopausal symptoms in breast cancer survivors: results of a randomized controlled trial. *J Natl Cancer Inst*. 2000;92:1054–1064.
62. Hermansen-Kobulnicky CJ. Symptom-monitoring behaviors of rural cancer patients and survivors. *Support Care Cancer*. 2009;17:617–626.
63. Hoekstra J, de Vos R, van Duijn NP, Schadé E, Bindels PJ. Using the symptom monitor in a randomized controlled trial: the effect on symptom prevalence and severity. *J Pain Symptom Manage*. 2006;31:22–30.
64. Kearney N, McCann L, Norrie J, et al. Evaluation of a mobile phone-based, advanced symptom management system (ASyMS) in the management of chemotherapy-related toxicity. *Support Care Cancer*. 2009;17:437–444.
65. Maunsell E, Brisson J, Deschenes L, Frasure-Smith N. Randomized trial of a psychologic distress screening program after breast cancer: effects on quality of life. *J Clin Oncol*. 1996;14:2747–2755.
66. McCall K, Keen J, Farrer K, et al. Perceptions of the use of a remote monitoring system in patients receiving palliative care at home. *Int J Palliat Nurs*. 2008;14:426–431.
67. Takeuchi EE, Keding A, Awad N, et al. Impact of patient-reported outcomes in oncology: a longitudinal analysis of patient-physician communication. *J Clin Oncol*. 2011;29:2910–2917.

68. Thewes B, Butow P, Stuart-Harris R, Greater Southern Area Health Service Screening Collaborative Group. Does routine psychological screening of newly diagnosed rural cancer patients lead to better patient outcomes? Results of a pilot study. *Aust J Rural Health*. 2009;17:298–304.
69. Velikova G, Keding A, Harley C, et al. Patients report improvements in continuity of care when quality of life assessments are used routinely in oncology practice: secondary outcomes of a randomised controlled trial. *Eur J Cancer*. 2010;46:2381–2388.
70. Berenson RA, Kaye DR. Grading a physician's value—the misapplication of performance measurement. *N Engl J Med*. 2013;369:2079–2081.
71. Ong WL, Schouwenburg MG, van Bommel ACM, et al. A standard set of value-based patient-centered outcomes for breast cancer: the International Consortium for Health Outcomes Measurement (ICHOM) Initiative. *JAMA Oncol*. 2017;3:677–685.

SUPPLEMENTAL MATERIAL

Supplementary Search Strategy: Literature Search Terms for Embase.

('breast tumor'/exp OR 'mastectomy'/exp OR (((breast* OR mamma*) NEAR/3 (cancer* OR tumor* OR tumour* OR carcino* OR neoplas* OR oncolog* OR malignan* OR resection* OR amputat*)) OR mammacarcinom* OR mastectom* OR mammectom* OR postmastectom*):ab,ti**) AND (**'patient-reported outcome'/de *OR 'patient reported outcome measure'/de* OR (('self report'/de) AND ('outcome assessment'/de OR 'quality of life'/exp OR 'quality of life assessment'/exp OR 'complication'/de OR 'symptom'/de OR 'wellbeing'/de OR 'psychological well-being'/de OR 'health status'/de)) OR (((patient* OR self) NEAR/3 (report* OR based OR centered OR centred OR rate OR rating OR rated OR questionnaire* OR assess* OR survey* OR index OR indices OR instrument*) NEAR/6 (outcome* OR measur* OR quality-of-life OR qol OR hrqol OR hrql OR ql OR symptom* OR complication* OR psychosocial OR 'well-being' OR wellbeing OR functioning OR disabil* OR 'health status')) OR ((patientreport* OR selfreport* OR patientcent*) NEXT/2 (outcome* OR factor* OR measur*)) OR PROM OR PROMs):ab,ti**) AND (**'health care quality'/de *OR 'quality of care'/de OR 'clinical effectiveness'/de OR 'clinical practice'/de* OR 'health care'/de *OR 'nursing care'/exp* OR (care OR ((clinical *OR nursing*) NEXT/2 (effectiveness* OR *practice**))):ab,ti**)**

Supplementary Table: Critical Appraisal Skills Programme (CASP) Quality Scoring According to the Study Design Performed.

Author	Study design	1	2	3	4	5	6	7	8	9	10	11	12
Abernethy et al.[31], 2009	Cohort	Yes	Yes	Yes	Yes	Yes	Yes	eTablets: - 94% easy to read - 99% easy to navigate - 90% comfortable - 89% easy to respond - 75% was satisfied with PCM reporting - 88% would recommend PCM - 74% felt PCM helped to report symptoms	Can't tell	Yes	Yes	Yes	N/A
Albert et al.[25], 2002	Cohort	Yes	Yes	Yes	Yes	Yes	Yes	100% found QoL profile a useful diagnostic tool, giving added value to both patients and doctors	Can't tell	Yes	Yes	Yes	N/A
Anderson et al.[32], 2015	RCT	Yes	Yes	Yes	No	Yes	Yes	- Pain severity: 0.6 vs 2.3 at time point 1, 1.2 vs 3.5 at time point 2 - Symptom severity: significant decrease between time point 1 and 2 - Symptom threshold: significant differences in ORs	P<0.01	No	Can't tell	Yes	N/A
Barbera et al.[40], 2015	Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Rate was 43% lower when screening with ESAS; for each additional ESAS assessment there was a 17% decrease rate of ED visits	Can't tell	Yes	Yes	Yes	N/A
Basch et al.[24], 2016	RCT	Yes	Yes	Yes	No	Yes	Yes	Effect size of 0.37 at 6 months	P<0.001	No	Can't tell	Yes	N/A
Berry et al.[49], 2014	RCT	Yes	Yes	Yes	No	Yes	Yes	SDS-15 score: -1.21	95% CI 0.2-2.20	Yes	Can't tell	Yes	N/A
Bock et al.[56], 2012	Cross-sectional	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes

Supplementary Table: Critical Appraisal Skills Programme (CASP) Quality Scoring According to the Study Design Performed. (continued)

Author	Study design	1	2	3	4	5	6	7	8	9	10	11	12
Børøsund et al.[39], 2014	RCT	Yes	Yes	Yes	No	Yes	Yes	Mean difference: - Symptom distress: 0.16 - Anxiety: 0.79 - Depression: 0.79	95% CI 0.06-0.25 95%CI 0.09-1.49 95%CI 0.09-1.49	No	Yes	Yes	N/A
Braeken et al.[30], 2013	RCT	Yes	Yes	Yes	No	Yes	Yes	Primary outcomes at 12 months: - Anxiety: 0.18 - Depression: 0.01 - Psychological distress: 0.46	P>0.05	Yes	Yes	Yes	N/A
Dean et al.[48], 2016	Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Breast reconstruction following mastectomy: - Satisfaction with breasts: 64.92 - Psychosocial wellbeing: 71.47 - Physical wellbeing 74.78 - Sexual wellbeing: 54.17	(61.92-67.92) (68.09-74.85) (77.43-77.13) (50.55-57.78)	Yes	Yes	Yes	N/A
Decker et al.[41], 2009	Cohort	Yes	Yes	Yes	Yes	Yes	Yes	The average sum of symptoms severity decreased by 4.35 in the intervention group	P= 0.21	Yes	Can't tell	Yes	N/A
Egbring et al.[20], 2016	RCT	Yes	Yes	Yes	Physicians (single-blinded)	Yes	Yes	Functional activity scores: from 90.385 (IQR 30.67) to 84.76 (IQR 18.29) (supervised group)	P=0.72	Yes	Can't tell	Yes	N/A
Fu et al.[45], 2016	Cross-sectional	Yes	Yes	Yes	Can't tell	Yes	Can't tell	Yes	Yes	Yes	Yes	Yes	Yes
Girgis et al.[26], 2009	RCT	Yes	Yes	Yes	No	Yes	Yes	Improves outcomes in the intervention group.	P<0.001 – P=0.01	Yes	Yes	Yes	N/A
Graf et al.[27], 2016	Cross-sectional	Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	No	Can't tell	Yes
Hahn et al.[37], 2004	Cohort	Yes	Yes	Yes	Yes	Can't tell	Yes	An electronic intervention with touch screen is useful in outcome measurement	Can't tell	Yes	Yes	Yes	N/A

Supplementary Table: Critical Appraisal Skills Programme (CASP) Quality Scoring According to the Study Design Performed. (continued)

Author	Study design	1	2	3	4	5	6	7	8	9	10	11	12
Holch et al.[21], 2017	Qualitative	Yes	Yes	Yes	Can't tell	Yes	Can't tell	Can't tell	Can't tell	Yes	Valuable	N/A	N/A
Javid et al.[46], 2016	Qualitative	No	Yes	Yes	Yes	Can't tell	Can't tell	Can't tell	No	Yes	Valuable	N/A	N/A
Judson et al.[35], 2013	RCT	Yes	Yes	Yes	No	N/A	Yes	Can't tell	Can't tell	Yes	Can't tell	Yes	N/A
Kelleher et al.[33], 2016	Observational	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Can't tell	Yes	Valuable	N/A	N/A
Kim et al.[44], 2016	Cohort	Yes	Yes	Yes	Yes	Can't tell	Yes	- Lower adherence (n=58) group reported 208/497 observations, while the higher adherence (n=20) group reported 289/497 observations - Self-report adherence was associated with an increase in the accuracy of depression screening	Precise	Yes	Yes	Can't tell	N/A
Knoerl et al.[19], 2017	Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Scores improves significantly due to the intervention	P=0.02	Yes	Yes	Yes	N/A
Knoerl et al.[22], 2017	Cohort	Yes	Yes	Yes	Yes	Yes	Yes	The intervention was feasible, with high usability, satisfaction and acceptability scores.	Precise	Yes	Yes	Yes	N/A
Kuijpers et al.[47], 2015	Qualitative	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Can't tell	Yes	Valuable	N/A	N/A
Leung et al.[38], 2016	Cohort	Yes	Yes	Yes	Yes	Can't tell	Yes	CAT scores were significantly correlated with the scores of both PROM questionnaires.	P<0.0001	Yes	Can't tell	Yes	N/A
Min et al.[36], 2014	Cohort	Yes	Yes	Yes	Yes	Can't tell	Yes	- Overall compliance rate 45% - Longitudinal compliance curve decreased from 100% (day 1) to 13.3% (day 90) - Unemployed women had a higher rate of compliance	Precise	Yes	Yes	Can't tell	N/A

Supplementary Table: Critical Appraisal Skills Programme (CASP) Quality Scoring According to the Study Design Performed. (continued)

Author	Study design	1	2	3	4	5	6	7	8	9	10	11	12
Mooney et al.[42], 2014	RCT	Yes	Yes	Yes	No	Yes	Yes	No significant difference in change of symptom severity between both groups (mean = 0.06)	P=0.58	Yes	Yes	Yes	N/A
Mooney et al.[43], 2017	RCT	Yes	Yes	Yes	No	Yes	Yes	Significantly less symptom severity in intervention group.	P<0.001	Yes	Yes	Yes	N/A
Ruland et al.[50], 2013	RCT	Yes	Yes	Yes	No	Yes	Yes	Can't tell	Can't tell	Yes	Can't tell	Yes	N/A
Snyder et al.[28], 2010	Cross-sectional	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No	No	Can't tell	Yes
Snyder et al.[11], 2013	Qualitative	Yes	Yes	Yes	Yes	Yes	Can't tell	Yes	No	Yes	Valuable	N/A	N/A
Stover et al.[23], 2015	Qualitative	Yes	Yes	Yes	Yes	Yes	Can't tell	Can't tell	Can't tell	Yes	Valuable	N/A	N/A
Wheelock et al.[34], 2015	RCT	Yes	Yes	Yes	No	Yes	Yes	More or new reported symptoms in intervention group (7.36 vs 3.2)	P=0.0045	Yes	Can't tell	Yes	N/A
Wu et al.[29], 2016	Qualitative	Yes	Yes	Yes	Yes	Yes	Can't tell	Yes	No	Yes	Valuable	N/A	N/A

Qualitative study:

[1]) Was there a clear statement of the aims of the research?
[2]) Is a qualitative/ quantitative methodology appropriate?
[3]) Was the research design appropriate to address the aims of the search?
[4]) Was the recruitment strategy appropriate to the aims of the research?
[5]) Was the data collected in a way that addressed the research issue?
[6]) Has the relationship between researcher and participants adequately considered?
[7]) Have ethical issues been taken into consideration?
[8]) Was the data analysis sufficiently rigorous?
[9]) Is there a clear statement of findings?
[10]) How valuable is the research?

Randomized controlled trial (RCT):

1) Did the trial address a clearly focused issue?
2) Was the assignment of patients to treatments randomized?
3) Were all of the patients who entered the trial properly accounted for at its conclusion?
4) Were patients, health workers and study personnel 'blind' to treatment?
5) Were the groups similar at the start of the trial?
6) Aside from the experimental intervention, were the groups treated equally?
7) How large was the treatment effect?
8) How precise was the estimate of the treatment effect?
9) Can the results be applied to the local population, or in your context?
10) Were all clinically important outcomes considered?
11) Are the benefits worth the harms and costs?

Supplementary Table: Critical Appraisal Skills Programme (CASP) Quality Scoring According to the Study Design Performed. (*continued*)

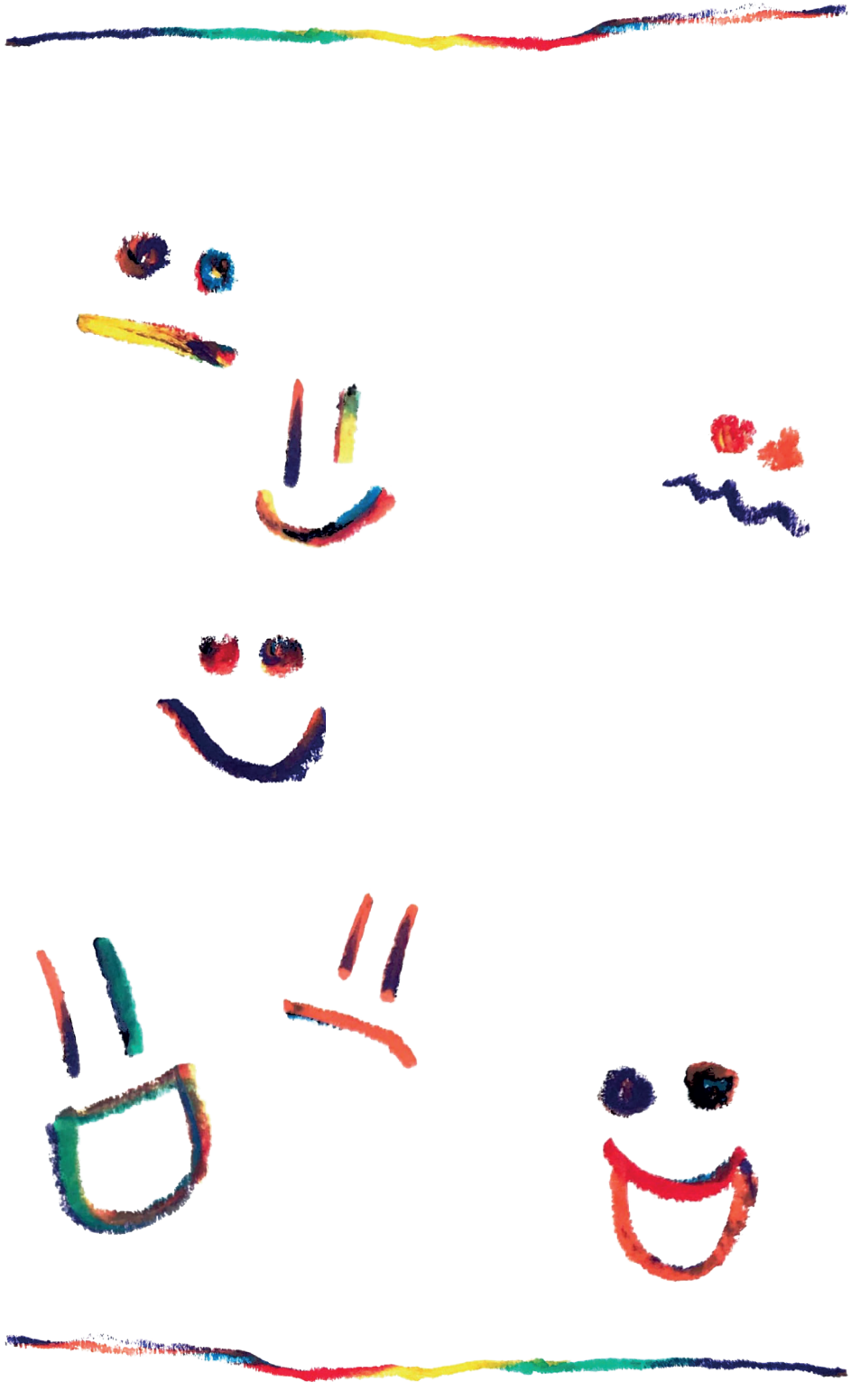
Author	Study design	1	2	3	4	5	6	7	8	9	10	11	12

Cohort study:

[1]) Did the study address a clearly focused issue?
[2]) Was the cohort recruited in an acceptable way?
[3]) Was the exposure accurately measured to minimize bias?
[4]) Was the outcome accurately measured to minimize bias?
[5]) A. Have the authors identified all important confounding factors?
B. Have they taken account of the confounding factors in the design and/or analysis?
[6]) A. Was the follow-up of subjects complete enough?
B. Was the follow-up of subjects long enough?
[7]) What are the results of this study?
[8]) How precise are the results?
[9]) Do you believe the results?
[10]) Can the results be applied to the local population?
[11]) Do the results of this study fit with other available evidence?
[12]) What are the implications of this study for practice?

Cross-sectional study:

1) Did the study address a clearly focused question / issue?
2) Is the research method (study design) appropriate for answering the research question?
3) Is the method of selection of the subjects (employees, teams, divisions, organizations) clearly described?
4) Could the way the sample was obtained introduce (selection)bias?
5) Was the sample of subjects representative with regard to the population to which the findings will be referred?
6) Was the sample size based on pre-study considerations of statistical power?
7) Was a satisfactory response rate achieved?
8) Are the measurements (questionnaires) likely to be valid and reliable?
9) Was the statistical significance assessed?
10) Are confidence intervals given for the main results?
11) Could there be confounding factors that haven't been accounted for?
12) Can the results be applied to your organization?



Chapter 6

A Prospective Analysis of Short-Term Patient-Reported Outcomes After Breast-Conserving Therapy and Mastectomy Without Reconstruction

Arvind Oemrawsingh, MD MHS[1]; Nikki van Leeuwen, PhD[1]; Marc A.M. Mureau, MD PhD[2]; Hester F. Lingsma, PhD[1]; Cornelis Verhoef, MD PhD[3]; Niek S. Klazinga, MD PhD[4]; Jan A. Hazelzet, MD PhD[1]; Linetta B. Koppert, MD PhD MHS[3]

[1] *Center for Medical Decision Making, Department of Public Health, Erasmus University Medical Center, Rotterdam, the Netherlands*
[2] *Department of Plastic and Reconstructive Surgery, Erasmus MC Cancer Institute, Erasmus University Medical Center, Rotterdam, the Netherlands*
[3] *Department of Surgical Oncology, Erasmus MC Cancer Institute, Erasmus University Medical Center, Rotterdam, the Netherlands*
[4] *Department of Public Health, Academic Medical Centre, University of Amsterdam, Amsterdam Public Health Research Institute, Amsterdam, the Netherlands*

Submitted.

ABSTRACT

Purpose

Although patient-reported outcomes (PROs) are increasingly assessed within breast cancer care, there is limited knowledge about short-term PRO scores following various surgical treatments. This study sought to investigate short-term PRO scores in breast cancer patients after either breast-conserving therapy (BCT) or conventional mastectomy.

Methods

Breast cancer patients, who had undergone surgical treatment at Erasmus MC between October 2015 and July 2019, were included. PROs were measured from prior to surgery to one year postoperatively using the BREAST-Q, EORTC-QLQ-C30 and EORTC-QLQ-BR23 questionnaires. Multiple imputation was used to account for missing responses. Within surgical treatment strategies, PRO domain scores at different follow-up points were compared with repeated measures ANOVAs.

Results

213 breast cancer patients were included: 143 (67%) underwent BCT and 70 (33%) conventional mastectomy. A significant improvement in "Body Image" at 12 months follow-up (Δ T0-T12 = +6.57, $p < 0.01$) was observed for patients following BCT, while "Physical Wellbeing" (Δ T0-T12 = –11.00, $p < 0.001$), "Physical Functioning" (Δ T0-T12 = –7.87, $p = 0.002$) and "Sexual Wellbeing" (Δ T0-T12 = –18.00, $p < 0.001$) significantly declined. Significantly reduced median "Satisfaction with Breasts" (Δ T0-T12 = –8.00, $p = 0.04$) and "Psychosocial Wellbeing" (Δ T0-T12 = –7.00, $p = 0.03$) scores were reported at one year following conventional mastectomy.

Conclusions

Most significant differences in PRO domain scores were observed in patients following BCT. The course of PROs during follow-up may help both patients and clinicians to better understand the impact of (surgical) treatment for breast cancer and manage expectations.

Keywords

Patient Reported Outcome Measure; Surveys and Questionnaires; EORTC-QLQ-C30; EORTC-QLQ-BR23; BREAST-Q; Breast Cancer; Value-Based Healthcare; Quality of Life

INTRODUCTION

In the Netherlands, approximately 16.400 women per year are diagnosed with breast cancer (14.000 with invasive and 2.400 with in-situ breast cancer).[1-3] Patients with early stage breast cancer show high survival rates[3], regardless of the type of breast surgery performed.[4-6] Considering that these oncological outcomes are comparable across multiple locoregional strategies (e.g. breast conserving therapy (BCT), conventional mastectomy, and mastectomy either followed by immediate or delayed breast reconstruction), there is an increasing desire for more knowledge about patient-reported outcomes (PROs) after breast cancer surgery.

PROs can be defined as patients' feedback on their own health condition including the effects of treatment, without external interpretation of healthcare providers.[7,8] They are, in addition to clinical (provider-reported) outcomes, increasingly being measured as part of the Value-Based Healthcare (VBHC) framework[9], to evaluate the quality of care delivered. The Erasmus University Medical Center (Rotterdam, the Netherlands) implemented the value-based breast cancer care strategy in October 2015 through a dedicated multidisciplinary breast cancer "disease team".[10,11] Essential components of this new strategy were the design of a "care pathway" with standardized time-points for follow-up consultations, and the routine collection of patient-reported outcome measures (PROMs) through an electronic data-collection tool linked to the electronic health record (EHR).

Although PROs are increasingly being assessed within breast cancer care[8] with an extensive body of literature on PROMs of breast cancer patients[12,13], there is still a substantial knowledge gap about the short-term trend of PRO domain scores following different breast cancer surgical strategies[14,15]. The aim of the present study was to gain knowledge on the short-term course of and correlations between common PRO scores in a breast cancer population undergoing either breast-conserving therapy (BCT) or conventional mastectomy.

METHODS

Study Population

Breast cancer patients, who had undergone surgical treatment between October 2015 and July 2019 at the Academic Breast Cancer Center of the Erasmus University Medical Center and had a follow-up of at least 12 months, were included in this study. The surgical treatments encompassed breast-conserving therapy (BCT)

and conventional mastectomy (without reconstruction). If patients previously had undergone BCT that was later followed by a mastectomy, they were categorized according to the final procedure.

Patients who had undergone breast reconstruction in the set study period were excluded from analysis, because many patients still had not finished the entire breast reconstruction trajectory (e.g., tissue expander exchange, additional corrections, nipple reconstruction) or had a completed breast reconstruction with insufficient follow-up. As the sample size of patients who had reached T12 follow-up within each breast cancer reconstruction strategy was very low, it was decided to include the reconstruction group in a future analysis when power and follow-up is sufficient.

Relevant patient-, tumor- and treatment-related characteristics were prospectively collected through the "Zorgmonitor", a web-based collection tool, and stored in a database. This observational study utilized this anonymized data on patient-reported outcomes as part of routine value-based breast cancer care (VBHC).[10] As per the locally approved institution's VBHC protocol, informed consent was not required from patients to store and use their anonymized information because the VBHC-strategy is considered standard of care in Erasmus University Medical Center.[10]

Patient-Reported Outcome Measures & Follow-Up

The selection of PROMs (e.g. BREAST-Q[16], EORTC-QLQ-Core 30[17], EORTC-QLQ Breast Cancer 23[18]) and collection time points (T0: prior to treatment; T6: six months after primary breast surgery; T12-T60: annually) were followed according to the International Consortium for Health Outcome Measurements' (ICHOM) Breast Cancer Outcome Set[19]. Patients received the web-based questionnaires by email. Weekly reminders up to three weeks were sent out by email, if the questionnaires remained incomplete. Nurse practitioners contacted patients by telephone after 4 weeks of no response and requested them to complete the questionnaires. Afterwards, patients' response was not actively pursued anymore.

For the current study, eleven subscales measured at T0, T6 and T12, were analyzed. These subscales included "Satisfaction with Breasts" "Psychosocial Wellbeing", "Physical Wellbeing", "Sexual Wellbeing" (BREAST-Q), "Cognitive Functioning", "Physical Functioning", "Emotional Functioning" (EORTC-QLQ-C30), "Body Image", "Breast Symptoms", "Arm Symptoms" and "Sexual Functioning" (EORTC-QLQ-BR23). The scoring and interpretation of these PROMs have been described in detail

elsewhere.[20,21] Briefly, all the subscales have a range from 0 – 100. A high score on a functional subscale indicates a healthy level of functioning, while a higher score on a symptom subscale signifies a higher degree of symptoms (burden).

Statistical Analysis

Data retrieval for the analyses was from a prospectively maintained electronic database. Continuous variables were expressed as a median with an interquartile range, and categorical variables were presented as counts with percentages. Continuous and categorical patient characteristics were compared between surgical treatment strategies (BCT vs. conventional mastectomy) and between respondents and non-respondents, using the Mann-Whitney U test and Chi-squared (χ^2) tests respectively.

Multiple imputation was used to account for missing patient-reported data within each surgical cohort by substituting a single alternative value for a missing PRO score at either T0, T6 and/ or T12. Imputation ensures the integrity of the cohorts by considering missing values as a source of variability to be averaged over.[22] Briefly, this method regressed an outcome on the patient characteristics and substituted the missing values with values from observed (non-missing) data that were closest to the model-predicted value for the respective PROM score. Age, BMI (body mass index), ASA (American Society of Anesthesiologists) classification, smoking, menopausal status, breast cancer gene mutation status, neo-adjuvant chemotherapy, surgical treatment strategy, family situation, pregnancy wish and employment status were the patient characteristics used as explanatory variables in the imputation model to create 10 imputed data sets. To account for uncertainty occurring due to imputation, final estimates were derived by pooling estimates from all imputed datasets. Subsequently, PRO scores between aforementioned follow-up points were compared with repeated measures ANOVAs (complete-case analysis) *within* surgical treatment strategies. The raw differences/ delta (Δ) in PRO scores between follow-up points were calculated.

Non-parametric Spearman's correlation coefficients were used to evaluate the overall correlation between individual domains across multiple PROMs. Correlations coefficients were estimated at baseline and 12-months follow-up within each surgical cohort. Correlations were specified as: 'very strong' ($.80 \leq r \leq 1.00$), 'moderately strong' ($.60 \leq r \leq 0.79$), 'fair' ($.30 \leq r \leq 0.59$) and 'poor' ($.00 \leq r \leq 0.29$).[23]

Two-sided p-values <0.05 were considered statistically significant. Statistical analyses were performed using RStudio version 3.6.1-© 2019 Inc. software, and

Statistical Package for Social Sciences (SPSS), Version 21.0 (IBM Corporation, Armonk, NY, USA).

RESULTS

Study Population

In total, 213 breast cancer patients, with a follow-up of at least 12 months, underwent surgery and completed PROMs between October 2015 and July 2019 (Table 1). 143 (67%) patients underwent BCT and 70 (33%) patients underwent conventional

Table 1 | Baseline Characteristics of All Breast Cancer Patients (N=213), According To Surgical Strategy.

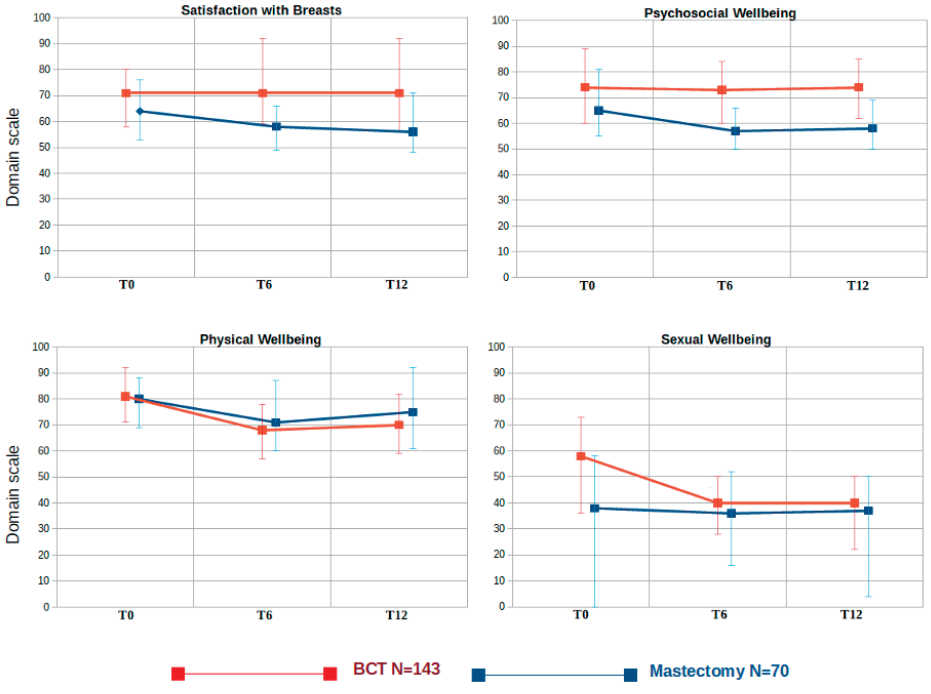
	All patients (N=213)	Missing/ Unknown, N (%)	BCT (N=143)	Conventional mastectomy (N=70)	p-value
Age, median [IQR]	58 [47 – 67]	3 (1.4)	58 [47.3 – 67]	59 [47 – 70]	0.57
Smoking (past/current), N (%)	23 (10.8)	3 (1.4)	17 (11.9)	6 (8.6)	0.35
BMI, N (%)		3 (1.4)			0.61
< 25	80 (37.6)		49 (34.3)	31 (44.3)	
25 – 30	96 (45.1)		68 (47.6)	28 (40)	
30 – 35	28 (13.1)		19 (13.3)	9 (12.9)	
> 35	6 (2.8)		4 (2.8)	2 (2.9)	
ASA-classification, N (%)		3 (1.4)			0.43
1	78 (36.6)		55 (38.5)	23 (32.9)	
2	118 (55.4)		77 (53.8)	41 (58.6)	
3	13 (6.1)		8 (5.6)	5 (7.1)	
4	1 (0.5)		-	1 (1.4)	
Menopausal status , N (%)		3 (1.4)			0.10
Pre-Menopause	49 (23)		31 (21.7)	18 (25.7)	
Peri-Menopause	31 (14.6)		26 (18.2)	5 (7.1)	
Post-Menopause	130 (61)		83 (58)	47 (67.1)	
Gene Mutation Carrier, N (%)	8 (3.8)	105 (49.3)	5 (3.5)	3 (4.3)	***0.02***
Family Situation, N (%)		3 (1.4)			0.42
Single	30 (14.1)		19 (13.3)	11 (15.7)	
Partner	70 (32.9)		50 (35)	20 (28.6)	
Partner + Child(ren)	98 (46)		65 (45.5)	33 (47.1)	
Child(ren)	12 (5.6)		6 (4.2)	6 (8.6)	
Pregnancy Wish, N (%)	22 (10.3)	106 (49.8)	15 (10.5)	7 (10)	0.67
Employment, N (%)	107 (50.2)	3 (1.4)	79 (55.2)	28 (40)	**0.03**

Note. BCT = breast conserving therapy; IQR = inter-quartile range; BMI = body mass index; ASA = American Society of Anesthesiologist score classification

mastectomy. Significantly more gene mutation carriers and unemployed women were in the conventional mastectomy group. Overall, the median follow-up time after the first surgery was 26 [IQR 18 – 32] months. A full overview of the types of surgeries performed is provided in the Appendix.

BREAST-Q Domain Scores within Surgical Strategies

At first glance, it is apparent that the trends of BREAST-Q scores, depicted by the curves, are similar across the two treatment strategies (Figures 1a-1d). Although median "Satisfaction with Breasts" scores did not change significantly over time for patients following BCT, they significantly kept declining following conventional mastectomy (T0-T12 difference = –8.00, p = 0.04). Median "Psychosocial Wellbeing" scores also significantly declined in patients following mastectomy (Δ T0-T12 = –7.00, p = 0.03). Median "Physical Wellbeing" scores significantly declined at T12 compared to baseline for patients who underwent BCT (Δ T0-T12 = –11.00, p < 0.001) and mastectomy (Δ T0-T12 = –6.00, p < 0.01). At one-year follow-up, "Sexual Wellbeing" significantly dropped only in patients who had undergone BCT (Δ T0-T12 = –18.00, p < 0.001).



Figures 1a-d | Median BREAST-Q Domain Scores Over Follow-Up within Surgical Treatment Strategies.

Note. Horizontal line: Median scores for all patients who had a follow-up of at least 12 months (N=213); Vertical line: inter-quartile range

EORTC-QLQ-C30 Domain Scores within Surgical Strategies

With regard to the EORTC-C30 subscales (Figures 2a-c), median "Physical Functioning" scores significantly declined for patients at 12 months following BCT (Δ T0-T12 = –7.87, $p < 0.01$). Median "Emotional Functioning" scores increased significantly over the one-year follow-up for patients who underwent BCT (Δ T0-T12 = +2.10, $p < 0.01$), surpassing baseline levels. At 12 months follow-up, "Cognitive Functioning" significantly declined for patients following BCT (Δ T0-T12 = –15.88, $p = 0.03$).

EORTC-QLQ-BR23 Domain Scores within Surgical Strategies

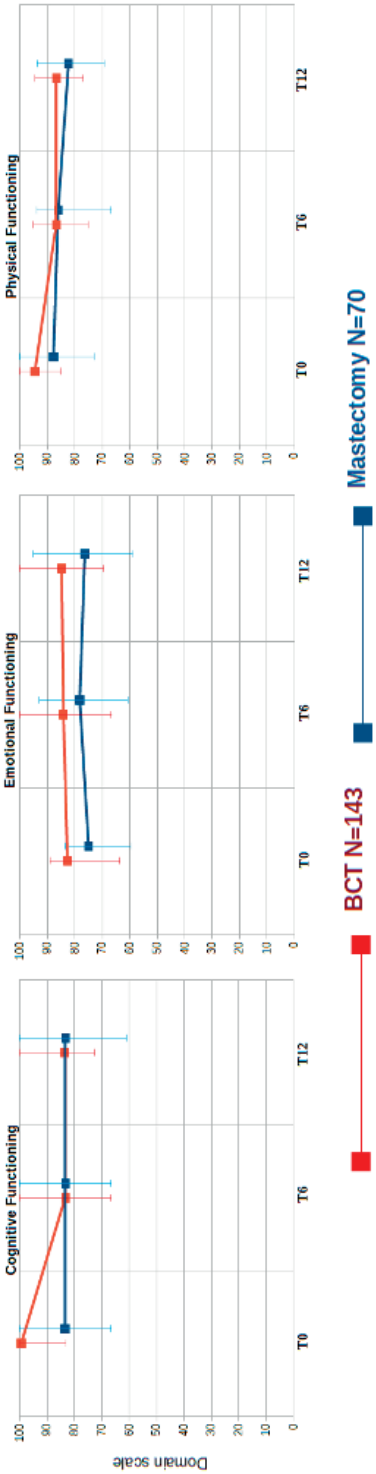
For patients who underwent BCT, there was a significant improvement in "Body Image" at 12 months follow-up (median Δ T0-T12 = +6.57, $p < 0.01$) (Figures 3a-d). Compared to baseline scores, both "Breast Symptoms" (median Δ T0-T12 = +7.22, $p < 0.01$) and "Arm Symptoms" (median Δ T0-T12 = +11.11, $p < 0.001$) were significantly higher at 12 months following BCT. Following conventional mastectomy, patients had increasing "Arm Symptoms" which was significantly higher at T12 compared to T0 (median Δ = +7.50, $p < 0.01$). Both "Breast Symptoms" and "Arm Symptoms", within each surgical cohort, were not significantly associated or correlated to sentinel lymph node biopsy or axillary lymph node dissection. There were no significant differences *within* each surgical cohort in reported "Sexual Functioning" over time.

Correlation across individual PROM domains

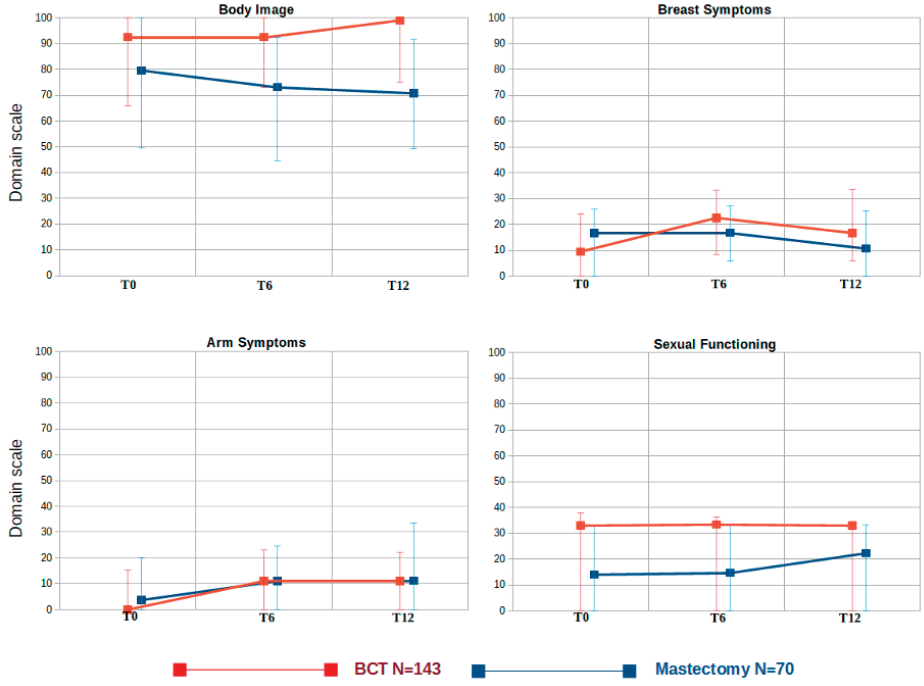
Overall, there was a moderately strong correlation between patient-reported "Physical Wellbeing" (BREAST-Q) and "Breast Symptoms" (EORTC-QLQ-BR23) at T12 following BCT ($r = –0.75$, $p < 0.001$) (Table 2). For patients who underwent conventional mastectomy, self-reported "Body Image" (EORTC-QLQ-BR23) was moderately correlated at T12 with "Satisfaction with Breasts" (BREAST-Q) ($r = 0.68$, $p < 0.001$) and "Psychosocial Wellbeing" (BREAST-Q) ($r = 0.71$, $p < 0.001$). Moderately strong correlations were also observed at T12 between self-reported "Physical Wellbeing" (BREAST-Q) and both "Breast Symptoms" ($r = –0.72$, $p < 0.001$) and "Arm Symptoms" (EORTC-QLQ-BR23) ($r = –0.61$, $p < 0.001$). At 12 months follow-up after mastectomy, there was a moderately strong correlation ($r = 0.64$, $p < 0.001$) between "Sexual Wellbeing" (BREAST-Q) and "Sexual Functioning" (EORTC-QLQ-C30).

Missingness Rates of PROMs Across Surgical Strategies

In the "original" (non-imputed) dataset only 118/213 (55%) patients completed the BREAST-Q questionnaires at *all three* follow-up points (baseline, T6 and T12).



Figures 2a-c | Median EORTC-QLQ-C30 Domain Scores Over Follow-Up within Surgical Treatment Strategies.
Note. Horizontal line: Median scores for all patients who had a follow-up of at least 12 months (N=213); Vertical line: inter-quartile range



Figures 3a-d | Median EORTC-QLQ-BR23 Domain Scores Over Follow-Up Per Surgical Treatment Strategy.

Note. Horizontal line: Median scores for all patients who had a follow-up of at least 12 months (N=213); Vertical line: inter-quartile range

Of the 213 participants, 124 (58%) and 124 (58%) completed the EORTC-C30 and EORTC-BR23 questionnaires respectively at *all three* follow-up points. Within the overall cohort, missingness was surprisingly higher at baseline than at 12 months follow-up for the EORTC-QLQ-C30 (23% vs. 21%) and the EORTC-QLQ-BR23 (26% vs. 17%). The BREAST-Q response rate declined from 71% at T0 to 67% at T12. No patient characteristics were significantly different between respondents and non-respondents at baseline across all three PROMs.

Table 2 | Pooled Spearman's Correlation Coefficients Between Individual PROM domains at T0 and T12 Per Surgical Strategy.

	Spearman's r
BCT at T0	
Physical Wellbeing – Physical functioning	0.34; p<0.001
Breast symptoms – Physical Wellbeing	–0.54; p<0.001
Arm symptoms – Physical Wellbeing	–0.46; p<0.001
Breast symptoms – Physical functioning	–0.31; p=0.001
Arm symptoms – Physical functioning	–0.39; p<0.001
Sexual functioning – Sexual Wellbeing	0.40; p<0.001
Satisfaction with Breasts – Body image	0.33; p<0.001
Satisfaction with Breasts – Breast symptoms	–0.14; NS
Satisfaction with Breasts – Emotional functioning	0.26; p=0.007
Psychosocial Wellbeing – Body image	0.42; p<0.001
Psychosocial Wellbeing – Emotional functioning	0.40; p<0.001
Psychosocial Wellbeing – Cognitive functioning	0.33; p<0.01
BCT at T12	
Physical Wellbeing – Physical functioning	0.40; p<0.001
Breast symptoms – Physical Wellbeing	**–0.75; p<0.001**
Arm symptoms – Physical Wellbeing	–0.59; p<0.001
Breast symptoms – Physical functioning	–0.40; p<0.001
Arm symptoms – Physical functioning	–0.36; p=0.001
Sexual functioning – Sexual Wellbeing	0.54; p<0.001
Satisfaction with Breasts – Body image	0.57; p<0.001
Satisfaction with Breasts – Breast symptoms	–0.35; p=0.002
Satisfaction with Breasts – Emotional functioning	0.20; p=0.04
Psychosocial Wellbeing – Body image	0.37; p=0.002
Psychosocial Wellbeing – Emotional functioning	0.22; NS
Psychosocial Wellbeing – Cognitive functioning	0.26; p=0.02
Conventional mastectomy at T0	
Physical Wellbeing – Physical functioning	0.24; NS
Breast symptoms – Physical Wellbeing	–0.43; p=0.001
Arm symptoms – Physical Wellbeing	–0.28; NS
Breast symptoms – Physical functioning	–0.43; p=0.002
Arm symptoms – Physical functioning	–0.33, p=0.008
Sexual functioning – Sexual Wellbeing	0.57; p<0.001
Satisfaction with Breasts – Body image	0.28; p=0.04
Satisfaction with Breasts – Breast symptoms	–0.19; NS
Satisfaction with Breasts – Emotional functioning	0.23; NS
Psychosocial Wellbeing – Body image	0.30; p=0.03
Psychosocial Wellbeing – Emotional functioning	0.47; p<0.001
Psychosocial Wellbeing – Cognitive functioning	0.36; p=0.01

Table 2 | Pooled Spearman's Correlation Coefficients Between Individual PROM domains at T0 and T12 Per Surgical Strategy. *(continued)*

	Spearman's r
Conventional mastectomy at T12	
Physical Wellbeing – Physical functioning	0.41; p=0.001
Breast symptoms – Physical Wellbeing	**–0.72; p<0.001**
Arm symptoms – Physical Wellbeing	**–0.61; p<0.001**
Breast symptoms – Physical functioning	–0.41; p=0.002
Arm symptoms – Physical functioning	–0.56; p<0.001
Sexual functioning – Sexual Wellbeing	**0.64; p<0.001**
Satisfaction with Breasts – Body image	**0.68; p<0.001**
Satisfaction with Breasts – Breast symptoms	–0.49; p=0.001
Satisfaction with Breasts – Emotional functioning	0.32, p=0.014
Psychosocial Wellbeing – Body image	**0.71; p<0.001**
Psychosocial Wellbeing – Emotional functioning	0.49; p<0.001
Psychosocial Wellbeing – Cognitive functioning	0.30; p=0.02

Note. BCT = Breast-Conserving Therapy; NS = not significant

DISCUSSION

Because long-term survival rates of breast cancer patients are still improving, it is imperative that more attention should be paid to long-term patient satisfaction and patient-reported breast-related quality of life outcomes.[24-26] By harnessing evidence-based knowledge on patients' short-term symptoms, satisfaction with breasts, and emotional, psychological and sexual wellbeing, patients can be better informed *after* the treatment decision.[27] Breast cancer-specific and generic PROMs were administered at specific follow-up times: baseline (T0), 6 months (T6), and 12 months (T12) after primary surgery, based on clinical meaning in order to focus on the impact of surgery while ensuring a feasible collection of the follow-up data. This prospective cohort study provides more knowledge about (health-related) quality of life issues of breast cancer patients before and after two surgical strategies, namely breast-conserving therapy (BCT) and conventional mastectomy.

For patients undergoing BCT, "Body Image" (EORTC-QLQ-BR23) did not significantly change during the initial 6 months after BCT but showed a significant improvement thereafter up to T12. This observation implies that BCT plus postoperative chemotherapy and/ or breast irradiation did not affect "Body Image" scales. Kindts et al.[15] reported a similar temporary impairment in this domain. Previous research found *long-term* "Body Image" scores to be higher in patients following BCT compared to mastectomy and autologous breast reconstruction.[26] Patient-

reported "Breast Symptoms" scores significantly increased at T6 following BCT, which subsequently declined but never reached baseline levels. The initial increase in breast-related complaints (e.g. breast edema, skin erythema, physical sensitivity, tenderness, pain) in BCT patients has been previously reported as well.[15,26] A significant increase was also seen in patient-reported "Arm Symptoms" domain scores in the 6 months following BCT, most likely attributed to either ALND (15% of study population) and/ or postoperative radiotherapy, which persisted at T12. Concurrently, both "Physical Wellbeing" and "Physical Functioning" significantly deteriorated over 6 months following BCT in the current study. These findings are very likely related to breast and arm morbidity due to ALND, or postoperative radiotherapy or chemotherapy. By discussing these observed PROM changes in preoperative consultations, expectations of future breast cancer patients can be better managed. For patients undergoing mastectomy without reconstruction, both "Satisfaction with Breasts" and "Psychosocial Wellbeing" scores significantly declined for patients at 12 months. This significant decrease in scores may be explained by the deleterious impact of conventional mastectomy plus possibly post-mastectomy irradiation, leading to a (unilateral) surgical defect, fibrosis and scar. Kouwenberg et al.[26] demonstrated that scores were the lowest following mastectomy compared to BCT and autologous breast reconstruction, suggesting the importance of breast preservation or reconstruction. To our knowledge, there are no other studies that have demonstrated the short-term changes in PROM scores after conventional mastectomy.[28] However, it is imperative that this knowledge gap be filled as patient-provider discussions about secondary breast reconstruction options can be enhanced with information on the trends of PROM scores after mastectomy and, ultimately, following reconstructive surgery. After all, previous research has shown a positive impact of delayed breast reconstruction on body image.[29]

The current study also provides new insights about correlations between individual domains of generic and breast-related PROMs. This secondary research question was inspired by ongoing efforts of organizations like the International Consortium for Health Outcomes Measurement (ICHOM) to harmonize the various PROMs in their Standard Disease Sets, in an attempt to reduce substantial registration burden on both patient and healthcare provider level. Overall, most correlations between the EORTC-QLQ-C30/BR23 and BREAST-Q were either "fair" ($r < 0.60$) or "poor" ($r < 0.30$), except between the BREAST-Q and EORTC-QLQ-BR23 at T12 for patients after conventional mastectomy ("Physical Wellbeing" with "Breast Symptoms" and "Arm Symptoms"; "Body Image" with "Satisfaction with Breasts" and "Psychosocial Wellbeing"; "Sexual Wellbeing" with "Sexual functioning"). These results indicate

that these three PROMs are largely complementary, with the exception of certain domains between the EORTC-QLQ-BR23 and BREAST-Q at T12. In a cohort study (N = 496), which included patients who underwent either BCT, conventional mastectomy, or mastectomy with (autologous/ implant) reconstruction, Lagendijk et al.[30] also reported the strongest correlations between "Body Image" with "Satisfaction with Breasts" (Pearson's r = 0.52) and "Psychosocial Wellbeing" (r = 0.66). A notable finding in the current study was the moderately strong correlation at T12 between "Breast Symptoms" (EORTC-QLQ-BR23) and "Physical Wellbeing" (BREAST-Q) for patients undergoing BCT. Interestingly, Stolpner et al.[31] tested the validity of the BREAST-Q Breast-Conserving Therapy Module against the EORTC-QLQ-BR23 and found a similar postoperative correlation coefficient ($r = -0.69$, $p < 0.01$) between these two domains, albeit questionnaires were administered one week after surgery. While the current study showed fair correlations between "Psychosocial Wellbeing" (BREAST-Q) and "Body Image" (EORTC-QLQ-BR23) following BCT, moderately strong correlations between aforementioned domains were observed for the mastectomy cohort. These findings suggest that the extent of surgery might enhance the impact that self-perception of body has on (postoperative) psychological wellbeing. If there is a desire to decrease the registration burden for patients by eliminating certain modules, future research is needed to corroborate the strong correlations between the EORTC-QLQ-BR23 and BREAST-Q.

Missingness of PROMs *at 12 months* was analyzed across the included questionnaires: BREAST-Q, EORTC-BR23 and EORTC-C30 domain scores were missing for 33%, 17%% and 21% of patients respectively. These low rates are were expected, considering the fact that longitudinal care has been associated with less PRO missingness when captured electronically.[25,32] A recently published systematic review on the implementation of PROMs in breast cancer care noted that pioneering clinical leadership can encourage the uptake of routine PRO measurements by both patients and care providers thus leading to less missing data.[25] Missing data can potentially lead to a decrease in power and precision and, more importantly, the possible introduction of bias in estimating of both between- and within-group effects (e.g. change over time).[33]

A considerable strength is the prospective nature of its data collection, which allows for a better insight in the trend of PROM scores over time, and how long-term PROM scores compare to their respective baseline level scores. In addition, the fact that analyses were performed on "real world" data (less restricted inclusion criteria and derived from multiple sources such as EHR and patient surveys) may extend the generalizability of the study results to a broader breast cancer popula-

tion[34]. However, this longitudinal study was not without limitations. First, due to the non-randomized design of this study, the PROM results per treatment strategy may have been partially confounded by treatment indications and by both known and unknown demographic (e.g. socio-economic status, educational background) and clinical variables (e.g. past history, tumor size, radiation treatment, systemic treatment) deliberately not controlled for in the analyses. Table 1 demonstrates some significantly different distributions of patient characteristics across the surgical cohorts. This is a major reason why the absolute median scores of PROMs in this study were not directly comparable *between* surgical strategies, and thus a causal relationship between surgical treatment and the trajectory of PROM scores could not be inferred[34]. Second, the relatively small sample sizes of particularly the mastectomy group compared to the BCT cohort may preclude the drawing of statistically powerful conclusions on the longitudinal data, particularly because of the skewed distribution of the BREAST-Q, EORTC-C30 and EORTC-BR23 scores. Finally, potential reasons for non-participation may have led to some selection bias. Patients who faced a language barrier or had a recurrence of breast cancer were not invited to fill in questionnaires. Also, patients who immediately after diagnosis were referred to the oncologist for neo-adjuvant systemic therapy, were regularly "missed" (at baseline) by the surgical nurse practitioners in the first period of the PROM initiative as they were solely being administered and monitored by the department of Surgical Oncology.

CONCLUSION

This observational study demonstrated a significant decline in patient-reported "Physical Wellbeing" and "Cognitive Functioning" scores following breast-conserving therapy, while patient-reported "Body Image" significantly increased at 12 months following BCT. "Satisfaction with Breasts" and "Psychosocial Wellbeing" scores significantly declined over one year after mastectomy. It is imperative to incorporate routine measurement of PROMs in clinical practice in order to equip patients with more information on the short- and long-term quality-of-life issues that can be expected following various breast cancer treatment strategies.

AUTHOR CONTRIBUTIONS

All authors have actively contributed to the work and meet the ICMJE criteria for authorship. Arvind Oemrawsingh, Nikki van Leeuwen, Marc Mureau, Jan Hazelzet,

Hester Lingsma and Linetta Koppert contributed to the conception and design of the study. Arvind Oemrawsingh and Linetta Koppert contributed to the acquisition of data. Arvind Oemrawsingh analyzed the data and drafted the initial manuscript. All authors critically revised the manuscript for intellectual content and approved the final manuscript.

FUNDING

Authors Arvind Oemrawsingh and Jan Hazelzet are supported by a grant from the Federation of Dutch University Medical Centers (NFU).

ACKNOWLEDGMENTS

The authors would especially like to thank breast cancer nurses and nurse specialists of the Academic Breast Cancer Center for their work during the implementation of the value-based breast cancer initiative.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

ETHICAL APPROVAL

Ethical approval was granted by the Medical Ethics Review Committee (MERC) of the Erasmus MC, Rotterdam, The Netherlands (MEC-2018-1015).

AVAILABILITY OF DATA AND MATERIAL

The anonymized dataset used and syntax for statistical analysis are available from the corresponding author upon reasonable request.

REFERENCES

1. https://www.rivm.nl/bevolkingsonderzoek-borstkanker.
2. https://www.cijfersoverkanker.nl/.
3. van der Waal D, Verbeek AL, den Heeten GJ, Ripping TM, Tjan-Heijnen VC, Broeders MJ. Breast cancer diagnosis and death in the Netherlands: a changing burden. *Eur J Public Health.* 2015;25(2):320-324.
4. Lagendijk M, van Maaren MC, Saadatmand S, et al. Breast conserving therapy and mastectomy revisited: Breast cancer-specific survival and the influence of prognostic factors in 129,692 patients. *Int J Cancer.* 2018;142(1):165-175.
5. Litiere S, Werutsky G, Fentiman IS, et al. Breast conserving therapy versus mastectomy for stage I-II breast cancer: 20 year follow-up of the EORTC 10801 phase 3 randomised trial. *Lancet Oncol.* 2012;13(4):412-419.
6. Poggi MM, Danforth DN, Sciuto LC, et al. Eighteen-year results in the treatment of early breast carcinoma with mastectomy versus breast conservation therapy: the National Cancer Institute Randomized Trial. *Cancer.* 2003;98(4):697-702.
7. Bredart A, Marrel A, Abetz-Webb L, Lasch K, Acquadro C. Interviewing to develop Patient-Reported Outcome (PRO) measures for clinical research: eliciting patients' experience. *Health Qual Life Outcomes.* 2014;12:15.
8. Howell D, Molloy S, Wilkinson K, et al. Patient-reported outcomes in routine cancer clinical practice: a scoping review of use, impact on health outcomes, and implementation factors. *Annals of oncology : official journal of the European Society for Medical Oncology.* 2015;26(9):1846-1858.
9. Porter MTE. Redefining Health Care: Creating Value-Based Competition on Results. Boston: Harvard Business School Press. 2006.
10. van Egdom LSE, Lagendijk M, van der Kemp MH, et al. Implementation of Value Based Breast Cancer Care. *European Journal of Surgical Oncology.* 2019;45(7):1163-1170.
11. Lagendijk M, van Egdom LSE, van Veen FEE, et al. Patient-Reported Outcome Measures May Add Value in Breast Cancer Surgery. *Annals of Surgical Oncology.* 2018;25(12):3563-3571.
12. Chen CM, Cano SJ, Klassen AF, et al. Measuring quality of life in oncologic breast surgery: a systematic review of patient-reported outcome measures. *Breast J.* 2010;16(6):587-597.
13. Kanatas A, Velikova G, Roe B, et al. Patient-reported outcomes in breast oncology: a review of validated outcome instruments. *Tumori.* 2012;98(6):678-688.
14. Batenburg MCT, Gregorowitsch ML, Maarse W, et al. Patient-reported cosmetic satisfaction and the long-term association with quality of life in irradiated breast cancer patients. *Breast Cancer Res Treat.* 2020;179(2):479-489.
15. Kindts I, Laenen A, van den Akker M, Weltens C. PROMs following breast-conserving therapy for breast cancer: results from a prospective longitudinal monocentric study. *Support Care Cancer.* 2019;27(11):4123-4132.
16. Pusic AL, Klassen AF, Scott AM, Klok JA, Cordeiro PG, Cano SJ. Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. *Plastic and reconstructive surgery.* 2009;124(2):345-353.
17. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *JNCI: Journal of the National Cancer Institute.* 1993;85(5):365-376.

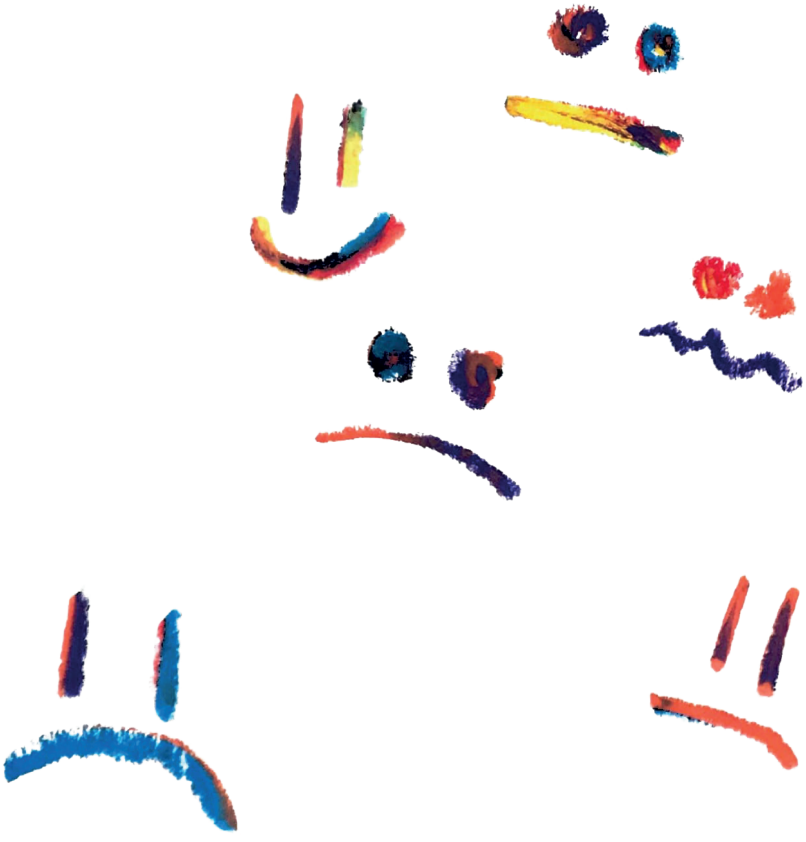
18. Sprangers MA, Groenvold M, Arraras JI, et al. The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: first results from a three-country field study. *Journal of Clinical Oncology.* 1996;14(10):2756-2768.
19. Ong WL, Schouwenburg MG, van Bommel ACM, et al. A Standard Set of Value-Based Patient-Centered Outcomes for Breast Cancer: The International Consortium for Health Outcomes Measurement (ICHOM) Initiative. *JAMA Oncology.* 2017;3(5):677-685.
20. https://www.eortc.org/app/uploads/sites/2/2018/02/SCmanual.pdf.
21. Fayanju OM, Mayo TL, Spinks TE, et al. Value-Based Breast Cancer Care: A Multidisciplinary Approach for Defining Patient-Centered Outcomes. *Ann Surg Oncol.* 2016;23(8):2385-2390.
22. Fielding S, Maclennan G, Cook JA, Ramsay CR. A review of RCTs in four medical journals to assess the use of imputation to overcome missing data in quality of life outcomes. *Trials.* 2008;9:51-51.
23. Chan YH. Biostatistics 104: correlational analysis. *Singapore Med J.* 2003;44(12):614-619.
24. Bodai BI, Tuso P. Breast cancer survivorship: a comprehensive review of long-term medical issues and lifestyle recommendations. *Perm J.* 2015;19(2):48-79.
25. van Egdom LSE, Oemrawsingh A, Verweij LM, et al. Implementing Patient-Reported Outcome Measures in Clinical Breast Cancer Care: A Systematic Review. *Value Health.* 2019;22(10):1197-1226.
26. Kouwenberg CAE, de Ligt KM, Kranenburg LW, et al. Long-Term Health-Related Quality of Life after Four Common Surgical Treatment Options for Breast Cancer and the Effect of Complications: A Retrospective Patient-Reported Survey among 1871 Patients. *Plast Reconstr Surg.* 2020;146(1):1-13.
27. Flanagan MR, Zabor EC, Romanoff A, et al. A Comparison of Patient-Reported Outcomes After Breast-Conserving Surgery and Mastectomy with Implant Breast Reconstruction. *Annals of Surgical Oncology.* 2019;26(10):3133-3140.
28. Cordova LZ, Hunter-Smith DJ, Rozen WM. Patient reported outcome measures (PROMs) following mastectomy with breast reconstruction or without reconstruction: a systematic review. *Gland Surg.* 2019;8(4):441-451.
29. Gopie JP, Mureau MA, Seynaeve C, et al. Body image issues after bilateral prophylactic mastectomy with breast reconstruction in healthy women at risk for hereditary breast cancer. *Familial cancer.* 2013;12(3):479-487.
30. Lagendijk M, van Egdom LSE, Richel C, et al. Patient reported outcome measures in breast cancer patients. *Eur J Surg Oncol.* 2018;44(7):963-968.
31. Stolpner I, Heil J, Feißt M, et al. Clinical Validation of the BREAST-Q Breast-Conserving Therapy Module. *Annals of surgical oncology.* 2019;26(9):2759-2767.
32. Nicholas GW, Steven PW, Gregory S, Amy PA, Thomas WL. Integration of Electronic Patient-Reported Outcomes Into Routine Cancer Care: An Analysis of Factors Affecting Data Completeness. *JCO Clinical Cancer Informatics.* 2017(1):1-10.
33. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Statistical Methods in Medical Research.* 2014;23(5):440-459.
34. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-World Evidence — What Is It and What Can It Tell Us? *New England Journal of Medicine.* 2016;375(23):2293-2297.

SUPPLEMENTAL MATERIAL

Supplementary Table: Overview of Breast Surgeries Performed on Study Population Between October 2015 – July 2019 with Follow-Up ≥ 12 months.

	All patients (N = 213)
Neo-Adjuvant Chemotherapy, N (%)	32 (15)
Breast Surgery Type, N (%)	
BCT	143 (67.1)
Mastectomy	70 (32.9)
Number of surgeries per patient, N (%)	
1	203 (95.3)
≥ 2	10 (4.7)
Breast Operated at Primary Surgery, N (%)	
Unilateral	203 (95.3)
Bilateral	10 (4.7)
	Patients Undergoing Primary Unilateral Breast Surgery (N = 203)
Unilateral Breast Procedure (Primary), N (%)	
BCT with J-wire localization	43 (21.2)
BCT with I-125	53 (26.1)
BCT by palpation	34 (16.8)
BCT (oncoplastic)	11 (5.4)
Mastectomy	62 (30.5)
Unilateral Axillary Procedure, N (%)	
SLNB	152 (74.9)
ALND	30 (14.8)
No treatment	21 (10.3)
	Patients Undergoing Bilateral Breast Surgery (N = 10)
Bilateral Breast Procedure (Primary), N (%)	
BCT with J-wire localization	1 (10)
BCT with I-125	1 (10)
Mastectomy	3 (30)
Other combinations	5 (50)
Bilateral Axillary Procedure (Primary), N (%)	
SLNB	2 (20)
ALND	-
No treatment	-
Other combinations	8 (80)

BCT = breast-conserving therapy; TE = tissue expander; LD = latissimus dorsi = DIEP: deep inferior epigastric perforator; SLNB = Sentinel lymph node biopsy; ALND = axillary lymph node dissection



Chapter 7 BREAST-Q Breast-Conserving Therapy Module: Normative Data From A Dutch Sample of 9059 Women

Arvind Oemrawsingh, MD MHS[1ǂ]; Marloes E. Clarijs, MD[2ǂ]; Andrea L. Pusic, MD MHS FACS FRCSC[3]; Hester F. Lingsma, PhD[1]; Cornelis Verhoef, MD PhD[2]; Jan A. Hazelzet, MD PhD[1]; Linetta B. Koppert, MD PhD MHS[2]

ǂ **Shared co-first authorship**

[1] *Center for Medical Decision Making, Department of Public Health, Erasmus University Medical Center, Rotterdam, the Netherlands*
[2] *Department of Surgical Oncology, Erasmus MC Cancer Institute, Erasmus University Medical Center, Rotterdam, the Netherlands*
[3] *Division of Plastic and Reconstructive Surgery, Patient Reported Outcomes, Value and Experience (PROVE) Center, Brigham and Women's Hospital, Massachusetts, USA*

Accepted in "Plastic and Reconstructive Surgery".

ABSTRACT

Background

The BREAST-Q, a patient-reported outcome measure for cosmetic and reconstructive breast surgery, is widely used in both clinical research and practice. The aim of this study was to acquire normative data of the BREAST-Q's Breast-Conserving Therapy module from a Dutch population sample and to compare it to existing normative BREAST-Q values.

Methods

Flyers with QR-codes, WhatsApp, and one academic center's Facebook and LinkedIn platforms were used to direct participants to self-complete an online version of four domains of the preoperative BREAST-Q Breast-Conserving Therapy module. BREAST-Q domain scores were log transformed to normalize the distribution. Univariable regression analyses were used to assess (non-linear) associations between age and BREAST-Q domain scores.

Results

Overall, 9059 questionnaire responses were analyzed. Median BREAST-Q domain scores were 64.0 ± SD 18.0 ("Satisfaction with Breasts"), 69.0 ± SD 21.0 ("Psychosocial Wellbeing"), 92.0 ± SD 20 ("Physical Wellbeing") and 59.0 ± SD 15.0 ("Sexual Wellbeing"). Age as a linear term was associated with log-transformed "Satisfaction with Breasts", "Psychosocial Wellbeing" and "Physical Wellbeing", while "Sexual Wellbeing" was a quadratic functions of age. Prior non-breast cancer-related surgery was a significant predictor for higher log-transformed "Satisfaction with Breasts" (β=0.04, $p<0.001$) and higher "Sexual Wellbeing" ($\beta = -0.05$, $p < 0.001$) scores. Compared to previously published normative data, small differences were found in mean BREAST-Q domain scores (mean differences ranging between 2.45–6.24).

Discussion

Normative Dutch BREAST-Q scores follow similar patterns across domains in comparison to previously published normative data. Normative Dutch BREAST-Q data enables future comparisons in breast-related satisfaction and quality of life issues of Dutch breast cancer patients against their age-matched peers.

Keywords

Patient-Reported Outcome Measure; Surveys and Questionnaires; BREAST-Q; Breast Cancer; Normative Data; Value-Based Healthcare

BACKGROUND

Approximately 17.000 women are diagnosed with breast cancer each year in the Netherlands of which 90% receive surgical treatment.[1,2] As the survival rates of (early-stage) breast cancer patients have considerably improved in the Netherlands[3,4], increasingly more attention is being paid on health-related quality of life (HRQoL) issues within this population.[5] Breast cancer patients often experience considerable burden related to numerous physical symptoms, psychosocial distress and impaired function during and after treatment.[6,7] Clinicians have increasingly become more aware of the importance of the patient perspective on surgical outcomes, with the goal to improve the quality of clinical breast cancer care.[8,9]

The BREAST-Q, published in 2009, is a well-validated, multi-scale and widely-used patient-reported outcome measure (PROM) for women undergoing breast surgery.[10] This condition-specific questionnaire measures the impact of oncological and/ or reconstructive breast surgery on different HRQoL domains, and has been used in a plethora of scientific publications in the last decade.[11] The use of PROM scores in breast cancer care may improve patient-provider communication[12], enhance shared-decision making, and manage current patients' expectations by informing them on the experiences of past patients[13]. In addition, as the use of PROMs in breast cancer clinical trials continues to rise, the availability of normative data may be helpful in enabling the interpretability and comparability of PROM scores across different trial arms.

However, it has become clear that normative scores are needed to fill the knowledge gap in the interpretation of PROMs, especially to provide context for the interpretation of patient scores following various treatment strategies. Normative data describes outcomes of a defined population without the specific condition of interest.[14] Currently, available normative BREAST-Q data has been published in two studies, both using samples of women of the United States with no previous history of breast cancer or breast surgery.[15,16] Therefore, the aim of this study was to collect and describe normative data of the BREAST-Q from a Dutch sample and compare it to internationally published normative values.

METHODS

Web-Based Questionnaire

An anonymous opt-in web-based questionnaire was developed in "LimeSurvey", a secure online survey tool provider.[17] This questionnaire contained the BREAST-Q and additional questions regarding age and breast cancer and/or breast operation in prior history. The questionnaire was prefaced with a study information sheet and a consent statement box to check. (Hyperlink: https://erasmusmcsurvey.erasmusmc.nl/mgz/ls/index.php/917249?lang=nl) The BREAST-Q©, developed at the Memorial Sloan-Kettering Cancer Center (NY, USA), is a surgery-specific instrument which assesses patient satisfaction and health-related quality of life in women undergoing different types of breast surgery.[10] Currently, BREAST-Q modules are Augmentation, Reduction/ Mastopexy and Breast Cancer Surgery, which includes scales for Mastectomy, Reconstruction, and Breast-Conserving Therapy (BCT) patients.[18]

The current study used the 2nd version (November 2017) of the preoperative scale of the BREAST-Q's Breast-Conserving Therapy module. This module containing 9 domains, of which 3 quality of life domains ("Physical Wellbeing" (10 items), "Psychosocial Wellbeing" (10 items), "Sexual Wellbeing" (6 items)) and 1 satisfaction domain ("Satisfaction with Breasts" (4 items)), was administered to participants in this study. Each scale is independent, and all items are scored on a Likert scale. According to the BREAST-Q protocol[18], the total score of each domain is separately transformed using a Rasch model to a number within the range of 0 – 100. A higher score means higher satisfaction or better health-related quality of life. The Dutch version was translated (both forward and backward translation) by a local academic prior to the conception and design of this study[18].

Recruitment of Study Participants and Data Collection

Three survey distribution techniques were chosen to reach as many potential respondents as possible. Firstly, digital and printed notifications with a QR-code hyperlink (to the web-based questionnaire) were distributed by medical students in the city center of Rotterdam for 6.5 days. Secondly, members of the research team disseminated the hyperlink to friends and family and invited them to share it with others. Thirdly, after a formal request was submitted, the Erasmus University Medical Center posted the survey link on its public Facebook and LinkedIn platforms. The status could be viewed and shared by people who are followers of the academic center's platforms. Participants were required to self-complete this questionnaire on the recruitment website of the Erasmus MC, after reading

the study information and giving informed consent. Participants were not compensated, and those who failed to complete the questionnaire did not receive a reminder. Data were collected from January till July 2020 and stored in the "LimeSurvey" database. Participants who did not finish or submit the questionnaire, who reported breast cancer in their past history, and who had discrepant responses (mastectomy reported in past history while no breast cancer reported), were not included in the analysis (Figure 1).

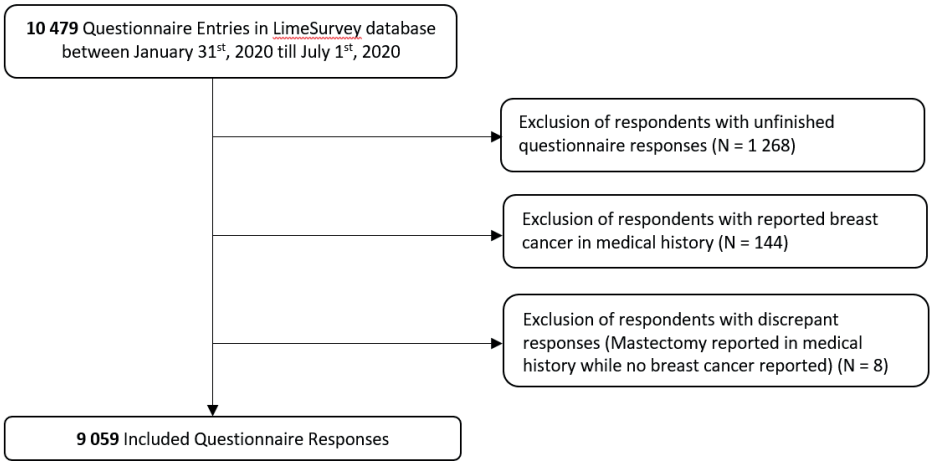


Figure 1 | Flowchart of Respondent Selection.

Statistical Analysis

Descriptive statistics, including medians and inter-quartile ranges (IQRs), were calculated to present the normative BREAST-Q data. The Mann-Whitney U test (non-parametric) was performed to compare BREAST-Q domain scores between women with and without non-breast cancer-related surgery in their past history. Skewness and kurtosis were calculated, with significant p values rejecting the null hypothesis for all domains. Therefore, natural log transformation was applied for all domain scores to normalize the distribution. Univariable models for each BREAST-Q domain were used to test for non-linearity of the effect of age by comparing models with age as a linear vs. age as a quadratic term. In addition, multivariable linear regression analysis was used to analyze whether there was a significant association between age, prior (non-breast cancer-related) breast surgery (e.g. surgery for fibroadenoma, cosmetic surgery) and (log-transformed) BREAST-Q domain scores. Cases that contained one or more missing items in a domain of the preoperative BREAST-Q version were excluded from analysis for that certain subscale. The t-test was used to compare means and standard devia-

tions of the current study's normative scores with previously published normative data. Two-sided p-values < 0.05 were considered statistically significant. Statistical analyses were performed using SPSS, Version 25.0 (IBM Corporation, Armonk, NY, USA).

Ethical Considerations

Formal approval from the local Medical Ethics Review Committee was not required, as the Dutch Medical Research (Human Subjects) Act did not apply to this study. The Legal Team and Department of Communications of the Erasmus University Medical Center approved to disseminate the survey link on the institution's social media platforms.

RESULTS

Study Participants

In total, 9 059 questionnaire entries were analyzed (Figure 1). Mean age for the overall group was 44 years ± SD 13, with most respondents representing the 40–49 and 50–59 age groups (Table 1). Eighty-one (0.9%) respondents did not complete the items of the "Sexual Wellbeing" domain. All other BREAST-Q domains were fully completed.

Table 1 | Characteristics of respondents (N = 9059), January – July 2020.

	N (%)
Age groups (years)	
18 – 29	1 385 (15.3)
30 – 39	1 776 (19.6)
40 – 49	2 482 (27.4)
50 – 59	2 424 (26.8)
60 – 69	837 (9.2)
≥ 70	155 (1.7)
Prior non-breast cancer-related surgery	
BCT (indication: Fibroadenoma)	22 (3.1)
Breast Reconstruction (Cosmetic)	86 (12.2)
Other	682 (96.7)

Dutch Normative BREAST-Q Scores

Overall, the median BREAST-Q domain scores were 64.0 ± SD 18.0 ("Satisfaction with Breasts"), 69.0 ± SD 21.0 ("Psychosocial Wellbeing"), 89.54 ± SD 12.48 ("Physical Wellbeing") and 60.38 ± SD 15.37 ("Sexual Wellbeing").

Median BREAST-Q domain scores stratified for age groups can be found in the Supplementary Material (Table 1). When comparing patients with (N = 705) versus without prior non-breast cancer-related surgery (N = 8 354), "Satisfaction with Breasts" was significantly higher (66.46 ± SD 20.26 vs. 64.05 ± SD 18.44, $p = 0.002$) while "Physical Wellbeing" was significantly lower (86.60 ± SD 14.26 vs. 89.79 ± SD 12.29, $p < 0.001$) for patients who had prior breast surgery.

Univariable Regression Analyses

Testing for linearity revealed age to be linearly associated with log-transformed "Satisfaction with Breasts" ($\beta = -0.001$, $p < 0.001$), "Psychosocial Wellbeing" ($\beta = 0.001$, $p < 0.001$) and "Physical Wellbeing" ($\beta = 0.001$, $p < 0.01$), while "Sexual Wellbeing" ($\beta = -3.7x10\text{-}5$, $p < 0.05$) was found to be a quadratic function of age (Figures 2a-2d). The model identified 34 years as the inflection point at which "Sexual Wellbeing" values began to decrease.

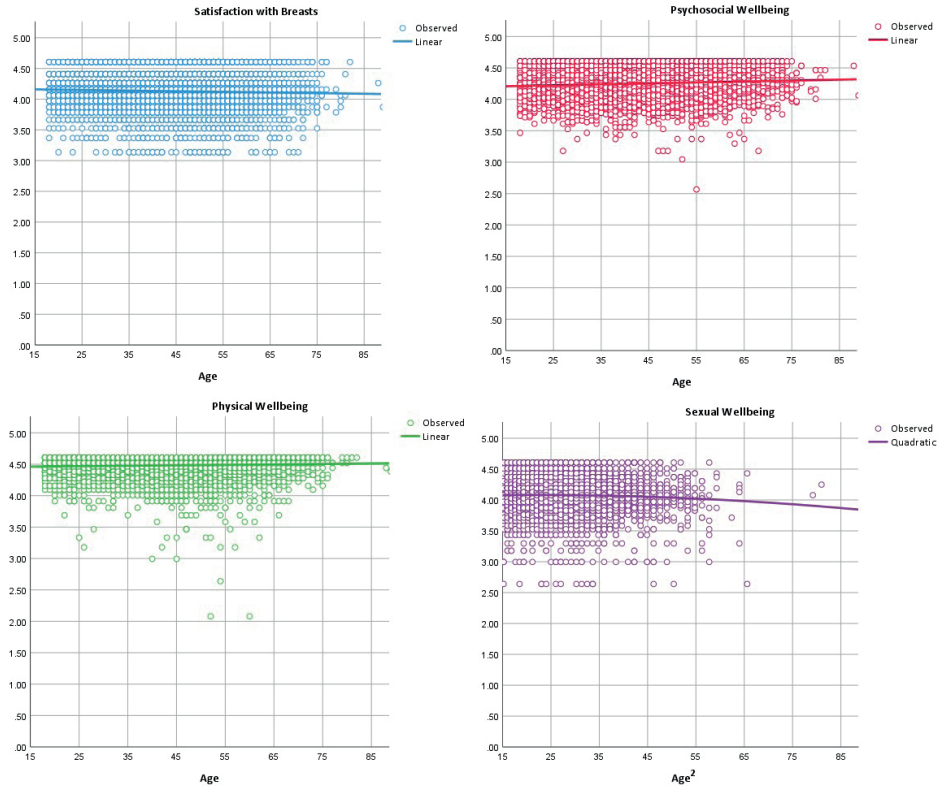
Multivariable Regression Analyses

Multivariable linear regression analyses (Table 2) confirmed age to be a significant predictor for log-transformed "Satisfaction with Breasts" ($\beta = -0.001$, $p < 0.001$), "Psychosocial Wellbeing" ($\beta = 0.001$, $p < 0.001$) and "Physical Wellbeing" ($\beta = 0.001$, $p < 0.001$). While prior non-breast cancer-related surgery was positively associated with log-transformed "Satisfaction with Breasts" ($\beta = 0.04$, $p < 0.001$), it had a negative association with "Physical Wellbeing" ($\beta = -0.05$, $p < 0.001$).

Comparison with Past Normative BREAST-Q Studies

Figure 3 demonstrates the variation in mean normative BREAST-Q domain scores between the current study and two internationally published studies by Mundy et al.[16] and Klifto et al.[15] on normative data of the BREAST-Q's Reconstruction module.

Normative scores for "Satisfaction with Breasts" were statistically higher in the current study compared to Mundy et al. (mean difference = 6.24, $p < 0.001$) and Klifto et al. (mean difference = 4.94, $p < 0.001$), as well as for "Sexual Wellbeing" compared to Mundy et al. (mean difference = 4.38, $p < 0.001$) and Klifto et al. (mean difference = 3.78, $p < 0.001$). A significant score difference for "Psychosocial Wellbeing" was found between the current study and Klifto et al. (mean difference =



Figures 2a-2d | Univariable Regression Plots Per Log-Transformed BREAST-Q Domain.

2.45; $p < 0.01$). Normative scores for "Physical Wellbeing" were lower compared to Mundy et al. (mean difference = 3.46, $p < 0.0001$), but higher compared to Klifto et al. (mean difference = 5.45, $p < 0.001$). In contrast to the current study ($\beta = 0.001$, p = NS), Klifto et al. found increasing age to be significantly associated with lower "Sexual Wellbeing" scores ($\beta = -0.3$, $p = 0.018$). Mundy et al. demonstrated younger age (< 40 years) to be associated with lower "Physical Wellbeing" scores, while the current study's cohort also revealed a minimal trend towards lower "Physical Wellbeing" scores in the younger age groups (Table 2).

Table 2 | Multi-Variable Linear Regression Coefficients For Log-Transformed BREAST-Q Domain Scores.

	Satisfaction with Breasts			Psychosocial Wellbeing			Physical Wellbeing			Sexual Wellbeing		
Adjusted R2	0.03			0.07			0.08			0		
	β	**SE**	**p-value**	**β**	**SE**	**p-value**	**β**	**SE**	**p-value**	**β**	**SE**	**p-value**
Age	–0.07	0.02	< 0.001	0.10	0.01	< 0.001	0.07	0.01	< 0.001	–0.02	0.01	NS
Prior non-breast cancer-related surgery	2.75	0.73	< 0.001	0.57	0.63	NS	–3.54	0.49	< 0.001	1.30	0.61	0.03

DISCUSSION

Because a breast cancer diagnosis has a significant impact on women both psychologically and physically[19], it is essential that there are baseline scores available representative of women's quality of life prior to a potential cancer diagnosis. The BREAST-Q preoperative module was originally considered as an accurate baseline; however, these patients are all aware of their underlying disease which immediately alters self-perception and quality of life.[15] According to normative data, one must take into account the degree of representation of the studied cohort. Previously published norm scores may not entirely reflect the normative scores for Dutch women due to cultural differences between American and Dutch female populations. Thus, the aim of this study was to collect and describe normative BREAST-Q scores from a Dutch sample, and to compare it to existing international normative data.

This study demonstrated through univariable regression analyses a negative linear relation between age and "Satisfaction with Breasts", while a positive association was observed between age and "Psychosocial Wellbeing". While some research has demonstrated higher self-reported body image and self-esteem in middle-aged women compared to younger women20 partially due to less self-objectification[21,22], other studies have concluded that age alone is not consistently associated to women's overall body image[23]. This study also identified prior non-cancer-related breast surgery to be positively associated with higher "Satisfaction with Breasts". These observations are in concordance with previous studies, which found improved breast satisfaction and quality of life following cosmetic breast surgery[24,25]. Unsurprisingly, prior breast surgery was associated to lower "Physical Wellbeing" scores, most likely due to prolonged complaints and physical discomfort (e.g. altered breast sensitivity, tenderness). A knowledge gap remains in how the strength between age and BREAST-Q domain scores alters when corrected for

other demographic variables. Ultimately, both age and non-breast cancer-related surgery do not seem to explain much of the variance in the BREAST-Q domain scores, as evident from the low R-squared values. The current study also compared its normative BREAST-Q domain scores to US scores published by Mundy et al.[16] and Klifto et al.[15] The differences between the current study's normative domain scores and those of the aforementioned studies could be due to both methodological differences and true inter-cultural differences between European and American females in terms of body image or satisfaction with breasts. The normative scores of Mundy et al. were not based on a random sample but, rather, on the response of 1201 participants from the Army of Women (AoW). The AoW is a strong community that is very aware of breast-related satisfaction and quality of life, possibly influencing responses and thus PROM scores. Klifto et al.'s normative values were obtained at multiple Johns Hopkins clinic sites by recruiting non-pregnant women without a prior history of breast cancer at their routine gynecology appointments. Despite the different data population characteristics and methodologies across the 3 studies (different BREAST-Q scales and different recruitment strategies), there was still a similar pattern observed across all BREAST-Q domains between the studies. This observation suggests that score differences are most likely attributable to different recruitment strategies, and that, contrary to expectations, inter-cultural differences in the perception of breast-related quality of life may have minimal influence on the normative scores. Utilizing country-specific normative scores may thus not be necessary, if the assumption is that similar normative scores would be obtained from other Western countries.

Klifto et al. showed a β of −0.3 for the "Sexual Wellbeing" domain in their regression analysis, meaning an increase in age per year results in a 0.3 point decrease in score, on a scale from 1 to 100. In the current study, an even smaller and non-significant β of 0.001 was observed. Although this result of Klifto et al. was significant, the question arises whether or in what extent such score differences on a scale from 1 to 100 are clinically important. It is thus not only important to take the statistical significance of PROM score differences between ages into consideration, but also the extent to which these differences are clinically meaningful. To date, there are very few publications on minimal clinically important differences (MCIDs) of the BREAST-Q[26,27], with varying definitions for MCID being used. MCIDs indicate the smallest change in PROM (domain) scores which patients perceive to be important or beneficial, and which would justify a change in patient management[27-29]. Voineskos et al.[27] determined, with distribution-based methods, a 4-point MCID on the BREAST-Q Reconstruction module (scale 0–100), based on 3052 patients who underwent breast reconstruction. It is important to

note that MCIDs are estimates that are subject to change, depending on the study population (e.g. age, BMI, socio-economic status, educational background, (cultural) setting) and the PROM ("outcome") of interest. Future research should focus more on calculating and validating MCIDs for the BREAST-Q, as interpretability is a cornerstone in utilizing PROMs for patient-centered care.

Strengths and Limitations

One of the key strengths of this study is the large size of the study sample. After the survey link was posted on the public Facebook and LinkedIn platforms of the Erasmus MC, a tremendous increase in respondents was observed. Because the questionnaire was widely available online and only in Dutch, the authors assume that respondents represent a heterogeneous Dutch population based cohort in terms of geographical distribution. Population-based studies are often based on probability samples of a reference population, therefore the exact sample strategy is important.[30] In this study, different recruitment methods were used to maximize the generalizability of the results. Compared to the population statistics in the Netherlands[31], both the age distribution and mean age of Dutch females (43.0 years in the population vs. 44 years in the study) were similar. Although this study successfully recruited a considerable number of respondents, it is possible that health-conscious women or women with a strongly positive/ negative body image were more likely to respond. The anonymity of the questionnaire may have partly remedied this effect but it may also have resulted in higher normative scores. The lack of socio-demographic characteristics (e.g. BMI, family status, educational background, socio-economic status) of the presented cohort prevents to state with utmost certainty that this current study is population-based.

In this sample, 6.8 percent reported a history of non-breast cancer related-surgery assuming this includes breast implants, augmentation, reduction and other cosmetic procedures. Some participants (N = 22) self-reported no previous history of breast cancer, but also reported having breast-conserving surgery in the past. This discrepancy may be partially explained by patients being treated for benign tumors such as fibroadenomas. As these cases also occur in the healthy female population, it still represents a normative cohort. Therefore, it was decided to only exclude patients who reported mastectomy but no history of breast cancer (N = 8). There is a possibility that this small number of patients has undergone preventive mastectomy, as approximately 40% of Dutch breast cancer gene mutation (BRCA 1/2) carriers undergo bilateral risk-reducing mastectomy (BRRM).[32] Prior non-breast cancer surgery for cosmetic purposes or medical reasons may affect the scores differently, for example in the "Satisfaction with Breasts" scale. To con-

clude, the possible heterogeneity of the prior non-breast cancer-related surgery group must be taken into account when interpreting the results.

A limitation of this study was the considerable sample size differences across some age groups, with older age groups (≥ 60 years) being less represented in this study. For Dutch women receiving the diagnosis of breast cancer, mean age is 57–62 years.[33,34] The mean age for the overall group in our sample was 44 years ± SD 13. This can be explained by the large number of respondents in the younger age groups, a potential consequence of using social media platforms. Older women may be less active on social media or may not have access to such platforms or a web-based questionnaire. Despite the fact that multiple recruitment methods were used, disproportionate sampling nevertheless emerged. However, the 40–49 and 50–59 age groups were represented by 2482 and 2424 participants respectively, still providing adequate power for the analyses. Survey data commonly exhibits disproportionate sampling, either as a result of bias in the sampling procedure or by design.[35] In this study, however, the sample size differences between age groups is most likely due to the latter. Another limitation were the limited questions that captured respondents' demographic profile. It was decided not to include questions on marital status, educational background, annual income and occupation, because authors were mindful of the fact that such "sensitive" questions can potentially affect survey outcomes by limiting the willingness of participants to complete the survey, the response rates to certain items, and the accuracy of responses.[36] As only two covariates (age and prior non-breast cancer-related surgery) were collected and used as predictors for model fitting, no variable selection procedure (e.g. forward or backward selection) was performed. A follow-up study will collect more covariates and proceed with a variable selection procedure to control for possible selection bias. Finally, as with most survey studies, some selection bias may have occurred due to self-selection. In this study, it was not possible to determine the response rate for returned questionnaires as a registration log was not kept for those women that viewed the survey hyperlink but chose not to open it. It does have to be noted that there were no significant differences in domain scores between patients who completed (N = 9059) and not fully completed (N = 1268) the questionnaire.

CONCLUSION

This study provides (age-dependent) normative values for the BREAST-Q, derived from a Dutch sample. These values may provide both clinicians and patients more

context when interpreting post-surgical BREAST-Q scores, thereby potentially managing patients' expectations and improving patient-provider communication in the consultation room. This study has demonstrated that age alone has a weak association with BREAST-Q domain scores on a scale of 0–100. Small differences were found between the current study's Dutch normative BREAST-Q scores and previously published US scores, while a similar pattern across domain scores was observed.

ACKNOWLEDGMENTS

- The authors would like to thank all women who participated in this study and so generously shared their experiences.
- Anonymized data and syntax of statistical analyses that support the study findings are available from the principal investigator on reasonable request.

CONFLICT OF INTERESTS:

- Authors Arvind Oemrawsingh and Jan Hazelzet are supported by a grant from the Netherlands Federation of University Medical Centers (NFU).
- Author Andrea Pusic is a co-developer of the BREAST-Q (owned by Memorial Sloan-Kettering Cancer Center) and may receive royalties when it is used in for-profit, industry-sponsored clinical trials.

REFERENCES

1. Integraal Kankercentrum Nederland. [Accessed July 2020]. Kerncijfers over borstkanker uit de Nederlandse Kankerregistratie 1989-2019. https://iknl.nl/getmedia/2f489dc8-a103-4fd6-89fe-e266e589366f/borstkanker-in-nederland_flyer_2020.pdf Accessed Accessed 10 July 2020.
2. Vondeling GT, Menezes GL, Dvortsin EP, et al. Burden of early, advanced and metastatic breast cancer in The Netherlands. *BMC Cancer.* 2018;18(1):262.
3. Janssen-Heijnen MLG, van Steenbergen LN, Voogd AC, et al. Small but significant excess mortality compared with the general population for long-term survivors of breast cancer in the Netherlands. *Annals of Oncology.* 2014;25(1):64-68.
4. van Maaren MC, de Munck L, de Bock GH, et al. 10 year survival after breast-conserving surgery plus radiotherapy compared with mastectomy in early breast cancer in the Netherlands: a population-based study. *The Lancet Oncology.* 2016;17(8):1158-1170.
5. Slatman J. The meaning of body experience evaluation in oncology. *Health Care Anal.* 2011;19(4):295-311.
6. Michael YL, Kawachi I, Berkman LF, Holmes MD, Colditz GA. The persistent impact of breast carcinoma on functional health status: prospective evidence from the Nurses' Health Study. *Cancer.* 2000;89(11):2176-2186.
7. Stein KD, Syrjala KL, Andrykowski MA. Physical and psychological long-term and late effects of cancer. *Cancer.* 2008;112(11 Suppl):2577-2592.
8. Kool M, van der Sijp JR, Kroep JR, et al. Importance of patient reported outcome measures versus clinical outcomes for breast cancer patients evaluation on quality of care. *Breast.* 2016;27:62-68.
9. Tevis SE, James TA, Kuerer HM, et al. Patient-Reported Outcomes for Breast Cancer. *Ann Surg Oncol.* 2018;25(10):2839-2845.
10. Pusic AL, Klassen AF, Scott AM, Klok JA, Cordeiro PG, Cano SJ. Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. *Plast Reconstr Surg.* 2009;124(2):345-353.
11. Cohen WA, Mundy LR, Ballard TN, et al. The BREAST-Q in surgical research: A review of the literature 2009-2015. *J Plast Reconstr Aesthet Surg.* 2016;69(2):149-162.
12. Galina V, Laura B, Adam BS, et al. Measuring Quality of Life in Routine Oncology Practice Improves Communication and Patient Well-Being: A Randomized Controlled Trial. *Journal of Clinical Oncology.* 2004;22(4):714-724.
13. Wright JG. Evaluating the outcome of treatment. Shouldn't We be asking patients if they are better? *J Clin Epidemiol.* 2000;53(6):549-553.
14. O'Connor PJ. Normative data: their definition, interpretation, and importance for primary care physicians. *Fam Med.* 1990;22(4):307-311.
15. Klifto KM, Aravind P, Major M, et al. Establishing Institution-Specific Normative Data for the BREAST-Q Reconstruction Module: A Prospective Study. *Aesthet Surg J.* 2020;40(6):NP348-NP355.
16. Mundy LR, Homa K, Klassen AF, Pusic AL, Kerrigan CL. Breast Cancer and Reconstruction: Normative Data for Interpreting the BREAST-Q. *Plast Reconstr Surg.* 2017;139(5):1046e-1055e.
17. LimeSurvey. https://www.limesurvey.org/. Accessed 01-07-2020.
18. BREAST-Q Version 2.0©. http://qportfolio.org/wp-content/uploads/2018/12/BREAST-Q-USERS-GUIDE.pdf. Accessed 01-07-2020.

19. Metcalfe KA, Semple J, Quan M-L, et al. Changes in Psychosocial Functioning 1 Year After Mastectomy Alone, Delayed Breast Reconstruction, or Immediate Breast Reconstruction. *Annals of Surgical Oncology.* 2012;19(1):233-241.
20. Borland H, Akram S. Age is no barrier to wanting to look good: women on body image, age and advertising. *Qualitative Market Research: An International Journal.* 2007.
21. Greenleaf C. Self-objectification among physically active women. *Sex roles.* 2005;52(1-2):51-62.
22. Tiggemann M, Lynch JE. Body image across the life span in adult women: the role of self-objectification. *Dev Psychol.* 2001;37(2):243-253.
23. Myers TA. Book Review: Body Image: A Handbook of Science, Practice, and Prevention. SAGE Publications Sage CA: Los Angeles, CA; 2012.
24. Coriddi M, Angelos T, Nadeau M, Bennett M, Taylor A. Analysis of satisfaction and well-being in the short follow-up from breast augmentation using the BREAST-Q, a validated survey instrument. *Aesthet Surg J.* 2013;33(2):245-251.
25. Dreher R, Blaya C, Tenório JLC, Saltz R, Ely PB, Ferrão YA. Quality of Life and Aesthetic Plastic Surgery: A Systematic Review and Meta-analysis. *Plast Reconstr Surg Glob Open.* 2016;4(9):e862-e862.
26. Cano SJ, Klassen AF, Scott A, Alderman A, Pusic AL. Interpreting clinical differences in BREAST-Q scores: minimal important difference. *Plast Reconstr Surg.* 2014;134(1):173e-175e.
27. Voineskos SH, Klassen AF, Cano SJ, Pusic AL, Gibbons CJ. Giving Meaning to Differences in BREAST-Q Scores: Minimal Important Difference for Breast Reconstruction Patients. *Plast Reconstr Surg.* 2020;145(1):11e-20e.
28. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol.* 2002;14(2):109-114.
29. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials.* 1989;10(4):407-415.
30. Szklo M. Population-based cohort studies. *Epidemiol Rev.* 1998;20(1):81-90.
31. https://www.cbs.nl/en-gb/visualisations/population-pyramid. Accessed Accessed 27-02-2021.
32. Heemskerk-Gerritsen BAM, Jager A, Koppert LB, et al. Survival after bilateral risk-reducing mastectomy in healthy BRCA1 and BRCA2 mutation carriers. *Breast cancer research and treatment.* 2019;177(3):723-733.
33. Johnson HM, Irish W, Muzaffar M, Vohra NA, Wong JH. Quantifying the relationship between age at diagnosis and breast cancer-specific mortality. *Breast cancer research and treatment.* 2019;177(3):713-722.
34. Lagendijk M, van Maaren MC, Saadatmand S, et al. Breast conserving therapy and mastectomy revisited: Breast cancer-specific survival and the influence of prognostic factors in 129,692 patients. *Int J Cancer.* 2018;142(1):165-175.
35. Maxwell A, Özmen M, Iezzi A, Richardson J. Deriving population norms for the AQoL-6D and AQoL-8D multi-attribute utility instruments from web-based data. *Qual Life Res.* 2016;25(12):3209-3219.
36. Tourangeau R, Yan T. Sensitive questions in surveys. *Psychological Bulletin.* 2007;133(5):859-883.

SUPPLEMENTAL MATERIAL

Table 1 | BREAST-Q Scores (Median with IQR) Stratified by Age Group.

Age groups (years)	N (%)	Satisfaction with Breasts	Psychosocial Wellbeing	Physical Wellbeing	Sexual Wellbeing
18 – 29	1 385 (15.3)	64 (53 – 77)	69 (60 – 80)	92 (80 – 100)	59 (50 – 66)
30 – 39	1 776 (19.6)	58 (53 – 71)	66 (60 – 80)	92 (80 – 100)	59 (50 – 66)
40 – 49	2 482 (27.4)	64 (53 – 71)	69 (60 – 83)	92 (80 – 100)	59 (50 – 66)
50 – 59	2 424 (26.8)	64 (53 – 71)	69 (62 – 87)	92 (85 – 100)	59 (50 – 70)
60 – 69	837 (9.2)	58 (48 – 71)	71 (62 – 87)	100 (85 – 100)	59 (48 – 66)
≥ 70	155 (1.7)	64 (53 – 71)	74 (64 – 87)	92 (85 – 100)	59 (48 – 67)
All patients	9 059	64 (53 – 71)	69 (60 – 83)	92 (80 – 100)	59 (50 – 66)*

Note. * Based on 8978 responses.

Chapter 8

Facilitators and Barriers For Implementing Patient-Reported Outcome Measures in Clinical Care: an Academic Center's Initial Experience

Marzyeh Amini, MSc[1ǂ]; **Arvind Oemrawsingh, MD MHS[1ǂ]**; Lisanne M. Verweij, PhD[2]; Hester F Lingsma, PhD[1]; Jan A. Hazelzet, MD PhD[1]; Frank Eijkenaar, PhD[3]; Nikki van Leeuwen, PhD[1]

ǂ **Shared co-first authorship**

[1] *Center for Medical Decision Making, Department of Public Health, Erasmus MC University Medical Center, Rotterdam, The Netherlands*
[2] *Erasmus MC University Medical Center, Rotterdam, The Netherlands*
[3] *Erasmus School of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands*

Health Policy. 2021 Jul 8:S0168-8510(21)00173-1.

ABSTRACT

Objectives

Although patient-reported outcome measures (PROMs) are increasingly used in clinical care, there is limited research studying the implementation of PROMs. The aim of this study was to explore the perspectives of healthcare providers and researchers in a large academic hospital on facilitators and barriers for implementing patient-reported outcome measures (PROMs) in clinical care.

Methods

A customized web-based questionnaire was developed and disseminated to healthcare providers and researchers across multiple medical departments involved in a value-based health care initiative in the hospital. Questionnaire statements were rated using a 5-point Likert scale ranging from "strongly agree" to "strongly disagree". In addition, 8 open-ended questions were included allowing respondents to mention additional facilitators and barriers for implementing PROMs. Descriptive statistics were used to summarize the results.

Results

In total, 61 participants from both surgical and non-surgical departments completed the survey. Most respondents (51%) were medical specialists and the median employment duration was 14 years. Frequently reported facilitators were the presence of a PROM coordinator in the (outpatient) clinic (85%), the integration of PROMs in the electronic health record (81%), and the intrinsic motivation of members involved in the implementation (N=9 open responses). Commonly reported barriers were language barriers (76%), IT-issues (N=17 open responses), and time constraints (N=14 open responses).

Conclusions

For the successful implementation of PROMs in clinical care, it is imperative that healthcare organizations consider accommodating motivated healthcare professionals, involving a PROMs coordinator, investing in an adequate IT-infrastructure, and the removal of language barriers.

Keywords

Patient-Reported Outcome Measures; Surveys and Questionnaires; Implementation; Value-Based Healthcare; Facilitators and Barriers, Health Services Research

INTRODUCTION

Patient-Reported Outcome Measures (PROMs) have increasingly become a topic of interest in both research and clinical practice[1], as part of an apparent shift in healthcare towards more value-based/value-driven systems[2,3]. PROMs can be defined as feedback directly from the patient on his or her health status (e.g. well-being, symptoms) and/or treatment, without external interpretation by healthcare professionals.[4,5] PROMs can serve multiple purposes including monitoring the development of individual patients' symptom burden and quality of life over the course of disease and/or treatment, facilitating patient-provider communication6, enhancing shared decision-making[7-10], examining quality of care[11], comparative effectiveness research[12], and their use in value-based payment[13].

In 2013, a value-based health care (VBHC) strategy was initiated within the Erasmus University Medical Center, a large academic healthcare institution in Rotterdam, the Netherlands. This VBHC initiative included the definition of standardized outcome sets (including both provider- and patient-reported outcomes), and the development of a collection tool to routinely capture these outcomes. Most of these standardized outcomes sets, which were developed by the International Consortium for Health Outcome Measurements (ICHOM), encompass both clinical (provider-reported) and patient-reported outcomes. The data collection tool was either linked to the electronic health record (EHR) or accessible through a web-based platform. Since the conclusion of the pilot phase of the initiative (2015-2019), the VBHC-concept has been successfully implemented for several conditions including breast cancer[14-16], cleft lip and palate[17], stroke[18], familial hypercholesterolemia[19] and Turner syndrome[20]. However, there have also been disease teams that did not (yet) succeed in implementing the collection of PROMs on a routine basis. Likewise, many hospitals and other healthcare organizations have had mixed success in implementing routine use of PROMs in a clinical setting with little insight in facilitating and impeding factors.[21]

Research aiming at obtaining a better understanding of healthcare professionals' experience or perspective on facilitators and barriers for implementing PROMs in clinical practice, is likely to be helpful to (other) hospitals or healthcare organizations that intend or are preparing to incorporate PROMs in clinical care. Therefore, the aim of this cross-sectional study is to identify the facilitators and barriers, as perceived by healthcare providers (e.g. physicians, nurses, paramedics) and researchers, to implementing PROMs in clinical care in a large academic medical center.

METHODS

Study Setting

Since 2013 as part of the aforementioned VBHC initiative, 38 multidisciplinary "disease teams" within the Erasmus University Medical Center have attempted to define PROMs and implement their routine capture (before patient visits) with review and feedback by treating healthcare professionals in the consultation room. The goal of the initiative was ultimately to provide care tailored to "what matters to patients" and to improve patient experiences of care by enhancing patient-provider communication and shared decision-making.[15] Following a pilot phase, ten of these teams (i.e. breast cancer, familial hypercholesterolemia, bladder cancer, stroke, brain tumors, otolaryngology disorders, cleft lip and palate, Turner syndrome, pediatric sickle cell anemia, adult sickle cell anemia) managed to successfully implement the concept in daily practice. The remaining 28 disease teams were in the earlier phases of implementation (see Supplementary Table 1), with some teams failing to complete the entire PROMs implementation process. These teams were intentionally also invited to participate in the current study in order to gain insights into the obstacles they experienced.

Study Design, Web-Based Questionnaire, Participants & Data Collection

According to Foster et al.[21], the implementation of PROMs can be divided into five stages:

1. Purpose and Intake (objectives and motivations for implementing PROMs)
2. Designing (making decisions on the process for implementing PROMs)
3. Preparing (investing time and resources to get an organization and its care providers ready to implement PROMs)
4. Commencing (the problems that ensue when organizations start to use PROMs)
5. Reflecting and Developing (evaluating the process of PROMs implementation and making adjustments)

For each stage, the facilitators and barriers reported in the systematic review by Foster et al.[21] provided the basic structure to develop the current study's questionnaire.

Healthcare professionals involved in the disease teams were invited to complete a voluntary, anonymous, opt-in, web-based questionnaire. Depending on the implementation stage the healthcare professionals were in during the study period, either the full version of the questionnaire or a selection thereof was sent to the respondents (see below). Reminders were sent two weeks after initial distribution.

The secure "LimeSurvey" online tool[22] was used to develop the questionnaire, which consisted of the following three sections:

1. Personal characteristics, including age, gender, professional role (e.g. doctor, nurse, researcher, physiotherapist), type of contract (e.g. part-time or full-time), department of employment, length of tenure at the Erasmus University Medical Center, and years of overall working experience within healthcare.
2. Eight open-ended questions on PROMs implementation within routine clinical practice, including perceived (dis)advantages, and perceived or experienced facilitators and barriers (Supplementary Table 2). This part was included in the questionnaire to give participants the opportunity to report their unique experiences that were not (fully) covered by the close-ended statements in the third section of the questionnaire. The open-ended questions were deliberately placed before the statements so respondents could not be influenced by them. Two questions were specifically related to facilitators and barriers, namely "What did you experience as the biggest factor promoting the implementation and/or use of PROMs in the (outpatient)clinic?" and "What did you experience as the biggest impeding factor when implementing and/or using PROMs in the (outpatient)clinic?". The responses to these two questions were analyzed and described in this study. The responses to the other six questions will not be discussed but are mentioned in Supplementary Table 2.
3. Statements about facilitators and barriers of PROMs implementation, with respondents being asked to rate their agreement on a five-point Likert scale (1 = strongly agree, 5 = strongly disagree). This part of the questionnaire is an adaptation of the previously validated "Barriers and Facilitators Assessment Instrument" (BFAI) developed by Peters et al. in the Netherlands.[23-25] The original 27-item BFAI instrument was developed to identify facilitators and barriers for implementation of preventive care (11 items) and innovations/guidelines (16 items), with facilitators and barriers being based on a literature study and an expert panel consensus procedure. The BFAI instrument allows for tailoring and addition of items.[24]

Additional statements in the current study's self-administered questionnaire were identified after literature review[8-10,21] by two authors (MA, AO) with added feedback from the other authors. The additional items were rephrased into statements and were then independently reviewed for relevance by two experts (one in VBHC and one in implementation research) within the Erasmus University Medical Center, and ultimately approved for inclusion by the research team.

According to Peters et al., possible facilitators or barriers can be related to the characteristics of the interventions, care providers, patients, and implementation context.[24] Therefore, items were divided in four domains to enable a categorized presentation of results:

1) Intervention characteristics (e.g. compatibility with existing routines, scientific evidence, consistency/ reliability)
2) Care provider characteristics (e.g. role perception/ attitude, lack of time, expertise/ specialty, relationship with patient, knowledge/ motivation)
3) Patient characteristics (e.g. age, ethnicity, language proficiency, health status)
4) Context characteristics (e.g. group standards/ socialization, presence of an information and administration system, setting).

A total of 37 statements (both positively and negatively phrased) were ultimately included in section three of the full version of the questionnaire. Since disease teams were in different phases of PROMs implementation and some facilitators or barriers were more (or even only) relevant for some phases, the number of statements included in the questionnaire was dependent on the implementation phases that respondents were currently in. Specifically, 28 statements were included in the questionnaire that was sent to participants that were in the "Designing" or "Preparing" phases, while the full questionnaire with all 37 statements was sent to participants that were already in the "Commencing" or "Reflecting and Developing" phases. Participants who were in the first phase ("Purpose and Intake") only responded to the questions in sections one and two of the questionnaire.

Although the accuracy and consistency of the questionnaire were not formally tested, it was pilot tested for comprehension among physicians and researchers (N=10). The results of this pilot indicated that completion of the whole questionnaire took approximately 15-20 minutes and that it was well-understood and positively received by the respondents. Minor textual amendments were made to the questionnaire following the pilot.

Data Analysis

Frequencies and percentages were computed to describe respondents' characteristics and responses to questions and statements. In reporting the results, we focused on statements that at least half of the respondents (dis)agreed or fully (dis) agreed with.

Data derived from the free-text responses to open-ended questions in the second section of the questionnaire were summarized independently by two authors (MA,

AO) and then compared. Exemplary quotes were selected to illustrate recurrent thematic elements and to provide more detailed context on experienced facilitators and barriers.

Analyses were performed with Statistical Package for Social Sciences (SPSS), Version 25.0 (IBM Corporation, Armonk, NY, USA).

Ethical Considerations

Ethical approval was not required for the anonymous web-based survey in this study as it was part of a quality-improvement project.[26] Informed consent of included participants was obtained, and their confidentiality was ensured.

RESULTS

Study Participants

In total, 61 of 187 invited healthcare providers and researchers completed the questionnaire (response rate = 33%) of which 1 respondent (2%) was in the "Purpose and Intake" phase, 23 respondents (38%) were in the "Designing" and "Preparing" phases, and 37 respondents (60%) were in the "Commencing" and "Reflecting and Developing" phases (Figure 1). The majority of respondents were

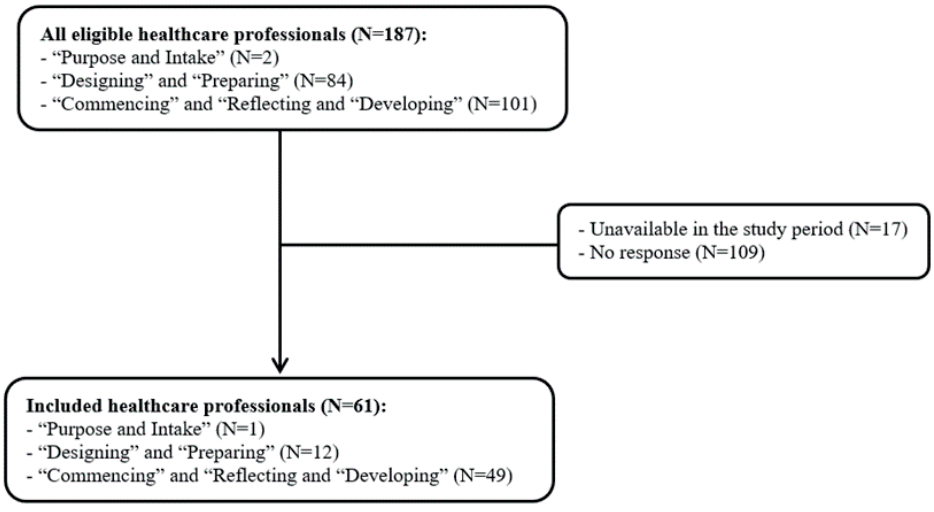


Figure 1 | Flowchart of Included Study Participants from Different Phases of PROM Implementation.

Note. The department of Quality & Patient Care (that offered support to disease teams during the implementation) identified eligible healthcare professionals, and provided information on the implementation phase that each disease team was in.

female (71%), medical specialists (51%), and non-surgical departments (56%). 17 out of 38 disease teams (44%) had at least one representative who completed the questionnaire (Table 1).

Table 1 | Demographic and Employment Characteristics of Study Participants (N=61).

Age, median [Min – Max]	46 [26 – 66]
Gender, N (%)	
Male	18 (29)
Female	43 (71)
Professional role, N (%)	
Medical specialist	31 (51)
Nurse	4 (7)
Nurse specialist	9 (15)
Paramedic	5 (8)
Physician researcher	2 (3)
Clinical psychologist	2 (3)
Physician assistant	2 (3)
Senior nurse consultant	2 (3)
Other roles	4 (7)
Department of employment, N (%)	
Surgical	27 (44)
Non-surgical	34 (56)
Conditions/diseases covered by respondents, N (%)	
Turner's syndrome	7 (11)
Cleft lip and palate	4 (7)
Gynecological cancers	4 (7)
Brain tumors	3 (5)
Head-and-neck cancer	3 (5)
Bladder cancer	3 (5)
Other	31 (50)
Not applicable (yet)	6 (10)
Work status, N (%)	
Full-time	30 (49)
Part-time	31 (51)
Employment duration in the hospital, median years [Min – Max]	14 [1 – 42]
Employment duration in healthcare, median years [Min – Max]	20 [3 – 45]
Status of implementation reached, N (%)	
"Purpose and Intake"	1 (2)
"Designing" and "Preparing"	23 (38)
"Commencing" and "Reflecting and Developing"	37 (60)

The respondents were mostly involved in the care for patients with Turner syndrome (N=7), cleft lip and palate (N=4), gynecological cancers (N=4), bladder cancer (N=3), brain tumors (N=3), and head-and-neck cancer (N=3). The median age of participants was 46 years old (min-max: 26–66). The median employment duration was 14 years (min-max: 1–42) in the Erasmus University Medical Center and 20 years (min-max: 3–45) in healthcare (Table 1).

Responses to Questionnaire Statements

At least half of the respondents (fully) agreed with eight of the positively phrased statements and (fully) disagreed with four of the negatively phrased statements (Table 2). These twelve factors are briefly described below. Most respondents (89%) perceived the purpose of implementing PROMs to be clear. In addition, 63% found the administered PROMs to be relevant for their patient population with an acceptable number of questions (56%). The majority of respondents (69%) also felt that open communication was possible within their disease team when providing feedback on the PROM implementation process. In addition, most participants (85%) perceived the presence of a coordinator in the (outpatient)clinic to be essential for implementing PROMs. Many respondents (59%) also reported that reviewing PROMs prior to the consultation and discussing them during the patient encounter, made the consultation more efficient. Furthermore, most respondents indicated that they were open to changing their old work routines (65%) and many respondents (88%) were confident that if they would use PROMs, they would be able to provide better care for patients. Moreover, they reported a willingness to use them in the future (76%). The majority (67%) of respondents were unafraid that the quality of their care might be assessed based on PROM results. Lastly, respondents (81%) indicated that the use of PROMs is or would be facilitated if they are integrated into the EHR, and most also reported a digital dashboard to be necessary for the consultation room with patients (79%).

On the other hand, the majority of respondents (76%) experienced a language barrier for patients with a primary language other than Dutch to be an impediment to the implementation of PROMs in the (outpatient) clinic. There were no other negatively phrased statements that at least half of the respondents (fully) agreed with or positively phrased statements that at least half of the respondents (fully) disagreed with (Table 2).

Table 2 | Responses to Statements on Facilitators and Barriers of PROM Implementation in Clinical Care.

		"Agree" and "Fully agree", N (%)	Neither "Agree" nor "Disagree", N (%)	"Disagree" and "Fully disagree", N (%)
Innovation characteristics				
1	The purpose of PROMs implementation in our care pathway is clear (+)	48 (89)*****	5 (8)	1 (2)
2	I found the care monitor platform user-friendly (+)	17 (32)	15 (28)	11 (20)
3	The layout of the care monitor platform makes it handy to use (+)	20 (37)	13 (24)	10 (19)
4	The current questionnaires are not relevant for my patient population (-)	8 (15)	4 (7)	34 (63)*
5	I think the PROMs questionnaires are too lengthy to fill in (-)	16 (29)	10 (19)	24 (44)
6	I think the PROMs questionnaires are too burdensome to fill in (-)	9 (17)	10 (19)	30 (56)*
7	I think the process of collecting PROMs is too extensive (-)	21 (39)	13 (24)	16 (30)
8	I am uncertain about the validity / sensitivity of PROMs to identify certain issues (for example, patients have not entered the PROMs correctly or have not reported symptoms, …) (-)	14 (26)	11 (20)	19 (35)
Care provider characteristics				
9	I feel that open communication is possible within the disease team to provide feedback on the process of implementing PROMs (+)	37 (69)*	9 (17)	3 (6)
10⊥	I am willing to use PROMs in the future (+)	29 (88)*	2 (6)	2 (6)
11⊥	I am confident that if I use PROMs, I can provide better care for my patients (+)	25 (76)*	3 (9)	2 (6)
12⊥	With the introduction of PROMs in the consultation room, the consumption of care (e.g. more referrals) is increased (+)	14 (43)	8 (24)	7 (21)
13	I have problems changing my old routines (-)	6 (11)	3 (6)	35 (65)*
14	I find it difficult to implement PROMs in (outpatient)clinical practice, because I have not been involved in setting up the PROMs implementation in our care pathway (-)	2 (4)	2 (4)	19 (35)
15	I find it difficult to implement PROMs in (outpatient) clinical practice, because I have not (yet) been trained to interpret PROMs and/ or discuss them with the patient (-)	7 (13)	5 (9)	23 (43)
16⊥	Managers/ directors do not cooperate in implementation of PROMs (-)	2 (6)	7 (21)	15 (46)
17⊥	Other colleagues do not cooperate in implementation of PROMs in outpatient clinic (-)	8 (24)	6 (18)	13 (39)
18⊥	I am afraid that the quality of our care might be assessed on the basis of PROMs results (-)	3 (9)	4 (12)	22 (67)*
Patient characteristics				
19	I notice that implementing PROMs in the (outpatient) clinic is difficult with patients with a language barrier (-)	41 (76)#	4 (7)	2 (4)
20	I notice that implementing PROMs in the (outpatient) clinic is difficult with patients with a non-Western cultural background (-)	13 (24)	14 (26)	15 (28)
21	I notice that implementing PROMs in the (outpatient) clinic is difficult with patients with a low socio-economic status (-)	15 (28)	13 (24)	16 (30)

Table 2 | Responses to Statements on Facilitators and Barriers of PROM Implementation in Clinical Care. (*continued*)

		"Agree" and "Fully agree", N (%)	Neither "Agree" nor "Disagree", N (%)	"Disagree" and "Fully disagree", N (%)
22	I notice that implementing PROMs in the (outpatient) clinic is difficult with seriously ill patients (-)	14 (26)	14 (26)	4 (7)
23	I notice that implementing PROMs in the (outpatient) clinic is difficult with elderly patients who are unable to complete a digital questionnaire (-)	26 (48)	4 (7)	2 (4)
24	I notice that implementing PROMs in the (outpatient) clinic is difficult with patients rarely visiting the (outpatient) clinic (-)	8 (15)	9 (17)	20 (37)
Context characteristics				
25	Discussing PROM results makes the consultation more efficient, because it is sooner clear in which domains the patient experiences any problems (+)	32 (59)*	6 (11)	9 (17)
26	The presence of a coordinator for the implementation of PROMs in the (outpatient) clinic is required (+)	46 (85)*	1 (2)	3 (6)
27	I could make better use of PROMs if they were integrated in the electronic patient file (+)	44 (81)*	2 (4)	1 (2)
28	There is sufficient time during the consultation to discuss PROMs results with patients (+)	16 (30)	8 (15)	22 (41)
29	I had enough time to be familiar with PROMs before it was implemented in our department (+)	25 (46)	5 (9)	10 (19)
30	The implementation of PROMs is sufficiently financially supported by the Erasmus University Medical Center (+)	16 (30)	10 (19)	13 (24)
31⊥	A digital platform / dashboard is necessary to facilitate the use of PROMs in the consultation room (+)	26 (79)*	2 (6)	3 (9)
32⊥	I think there is sufficient support with the analysis and interpretation of PROMs (+)	10 (30)	9 (27)	11 (33)
33	I find it difficult to implement PROMs in (outpatient) clinical practice, because there is insufficient supportive staff (-)	20 (37)	11 (20)	10 (19)
34	I find it difficult to implement PROMs in (outpatient) clinical practice, because IT-staff insufficiently supports the implementation of PROMs (-)	12 (22)	10 (19)	23 (43)
35	I find it difficult to implement PROMs in (outpatient) clinical practice, because patients do not cooperate in the collection of PROMs (-)	2 (4)	11 (20)	27 (49)
36	I find it difficult to implement PROMs in (outpatient) clinical practice, because measurement of PROMs is too time consuming (-)	14 (26)	10 (19)	21 (39)
37	Implementation of PROMs has technical problems (e.g. accessing the electronic questionnaire, visualizing the results, …) (-)	22 (41)	8 (15)	11 (20)

Notes. (+) indicates positively phrased statement and (-) indicates negatively phrased statement.
Differences between responses in the table and total number of participants are due to the response "I do not know" or no response at all (considered as missing value). Across items, this varied between 11% and 62%.
* Identified as a facilitator
# Identified as a barrier
⊥ N = 33. For all other items, N=54. The difference is related to the number of participants in different phases of PROMs implementation.

Responses to Open-Ended Questions

All 61 study participants responded to the open-ended question "What did you experience as the biggest facilitating factor in the implementation of PROMs in the (outpatient) clinic?" (Table 3). Nine respondents (15%) reported that "motivation/enthusiasm/interest within the disease team" is needed for successfully implementing PROMs. This factor was the only reported facilitator by employees across most implementation stages, namely "Designing", "Preparing", "Commencing" and "Reflecting and Developing". Some participants (N=4) noted "collaboration with multiple disciplines on every level of care" as an encouraging factor. A "more structured patient-provider communication about PROM results" was another reported facilitating factor (N=6). Three respondents mentioned that "sufficient support from the central VBHC team" promotes the implementation. Other reported facilitators (N=2 each) were "well-organized IT", "the provision of a guideline for interpreting PROMs", "the past experience of other colleagues", "the team-building around one clinical condition", "sufficient resources", and "the presence of a nurse practitioner".

All respondents also responded to the open-ended question "What did you experience as the biggest impeding factor in the implementation of PROMs in the (outpatient) clinic?" (Table 3). Seventeen respondents indicated issues with the IT infrastructure to be impeding factors in the implementation process. These issues included "access to the web-based platform" (N=8), the "visualization of the dashboard" (N=1), the "response time of the IT-platform" (N=3), and "integration of the PROMs into the EHR" (N=5). In addition, "time constraints of care providers" in combination with the "labor-intensive nature of the incorporation of PROMs in care" were frequently reported (N=14). These two barriers (i.e. IT-related challenges and time constraints/ labor intensity) were most reported across most phases, including "Designing", "Preparing", "Commencing" and "Reflecting and Developing". Furthermore, "insufficient staff for support" and "insufficient staff for coordination" were also reported as barriers (N=10). Other experienced barriers were "challenges in multidisciplinary collaboration due to conflicting interests, schedules, and task division" (N=3), "lack of uniformity" (N=3), and "excessive duration of the implementation process" (N=2). Finally, barriers reported by single respondents were "insufficient information", "language barrier for patients with a primary language other than Dutch", "inappropriate organization of care", "learning new work routines with PROMs", and "different motivation levels within the disease team to use PROMs".

Table 3 | Responses to Selected Open-Ended Questions, Stratified by Stages of PROM Implementation.

Questions	Phases of implementation	Summary of respondent comments
What did you experience as the biggest facilitating factor in the implementation and / or use of PROMs in the (outpatient) clinic?	"Purpose & Intake" (N=1)	- No response/Not applicable (N=1)
	"Designing" & "Preparing" (N=23)	- Intrinsic motivation/ enthusiasm/ interest (N=4) - Collaboration with multiple disciplines (N=4) - Support from the VBHC-team (N=3) - Better insight into desires and problems of patients (N=2) - Other (N=4): Communication about PROMs Input and appreciation of patients Guideline for PROM interpretation Tailored care - No response/Not applicable (N=6)
	"Commencing" & "Reflecting & Developing" (N=37)	- More structured/ efficient patient-provider communication (N=6) - Involvement of patients in own care (N=5) - Motivation/ enthusiasm/ effort of the "disease team" (N=5) - Better preparedness/ more information on how patients are doing prior to consultation (N=4) - Presence of a nurse practitioner (N=2) - Well-arranged IT-system (N=2) - Review of (the trend) of outcomes (N=2) - Improved/ more efficient patient care (N=2) - Other (N=4): Past experience of other colleagues Full support (of "disease team" members) on the introduction of PROMs Teambuilding around one disease Presence of a champion for PROMs Better collaboration on every level of care Regional agreements on collaboration with partners (e.g. insurance companies) Providing enough resources in outpatient clinic - No response/Not applicable (N=5)

Table 3 | Responses to Selected Open-Ended Questions, Stratified by Stages of PROM Implementation. (*continued*)

Questions	Phases of implementation	Summary of respondent comments
What did you experience as the biggest impeding factor in the implementation and / or use of PROMs in the (outpatient) clinic?	"Purpose & Intake" (N=1)	- Lack of support (N=1)
	"Designing" & "Preparing" (N=23)	- Time constraints/ labor-intensive (N=6) - IT-issues (e.g. web-based platform not integrated into the EHR) (N=4) - Diversity/ lack of uniformity in implementation process (N=3) - Challenges in multidisciplinary cooperation (e.g. conflicting interests, schedules, task division) (N=3) - Lack of support (N=2) - Other (N=2): Absence of a nurse practitioner Lack of enough information - No barriers experienced (N=2) - No response/Not applicable (N=5)
	"Commencing" & "Reflecting & Developing" (N=37)	- IT-issues (e.g. difficult access to web-based platform, platform not integrated into the EHR, slow working speed of the platform, visually unattractive dashboard, unregistered email addresses at baseline, inflexible automatic distribution system) (N=13) - Time constraints/ labor-intensive (N=8) - Insufficient outpatient clinic staff/ nurse practitioner for coordination (N=4) - Lack of support (N=4) - Excessive duration of the implementation (N=2) - Communication within the "disease" team (N=2) - Other (N=6): Language barrier in patients Non-response of patients Inappropriate organization of care Learning new routines in using questionnaires Different levels of motivation/activity to use PROMs among "disease team" members - No response/Not applicable (N=4)

Note. EHR = electronic health record; PROMs = Patient-Reported Outcome Measures; VBHC = Value-Based Healthcare

DISCUSSION

Although the adoption of PROMs in routine clinical care has been widely advocated, few studies have analyzed facilitators and barriers to the implementation of PROMs in clinical practice as perceived or experienced by involved healthcare professionals.[21] This cross-sectional study sought to investigate the views of healthcare professionals on the implementation of PROMs in clinical care in the context of a VBHC-initiative in an academic medical center. Overall, our findings highlight that for a successful implementation of PROMs in clinical practice, it is essential for healthcare organizations to support motivated healthcare professionals, to arrange for a PROM coordinator in the (outpatient) clinic, to invest in an adequate IT-infrastructure, and to remove language barriers for non-Dutch speaking patients by making multilingual surveys available on an interface. Many factors that influence healthcare professionals' ability and desire to implement the routine measurement of PROMs seem to be bi-directional: they can act as either facilitators or barriers.

Respondents had mainly positive attitudes towards implementing PROMs in clinical practice, with most reporting its purpose to be clear and the utilized questionnaires to be relevant to their patient population and not too burdensome to fill in. The majority of study respondents expressed confidence in PROMs contributing to better care, with most respondents reporting being open to changing old work routines and willing to use PROMs in the future. This is in line with a study by Boyce et al., who found that "healthcare professionals value PROMs when they are useful for the clinical decision-making process"[9]. In the current climate of skepticism on the validity and interpretability of PROMs[27], it is crucial that prior to implementation, consensus is reached among all involved healthcare professionals on the goal of routine PRO measurement (e.g. detecting previously unrecognized health issues[28], monitoring disease progression[29], stimulating better communication[30] and promoting shared decision-making[31]).

The responses to both questionnaire statements and open-ended questions demonstrated that respondents experienced open communication about the PROM implementation process within their disease team. This seems related to the factor "intrinsic motivation/enthusiasm/interest", which was the only factor that was reported as a facilitator across almost all implementation stages. Motivation has been associated with encouraging teamwork and creating a sense of solidarity between colleagues.[32] This essential component can enhance collaboration between healthcare professionals in a multidisciplinary team, which does not

always turn out to be easy. In this study, "multidisciplinary collaboration" within a disease team was also mentioned as a facilitator, as was "teambuilding around one disease/condition". Correspondingly, "different motivation levels within the disease team to use PROMs" and "challenging multidisciplinary cooperation due to conflicting interests, schedules, and task division" were described as barriers. Healthcare professionals need to learn to work together in a multidisciplinary way, by exploring the boundaries of each other's disciplines and being open to other insights, raise mutual trust, and create a sense of joint responsibility.[33]

Other interesting facilitators reported by healthcare professionals were (the prospect of) "better preparedness with more information on how patients are doing prior to consultation" and (the prospect of) a "more structured/ efficient patient-provider communication". In addition to these factors, the "involvement of patients in their own care" was also reported as a facilitator. These factors suggest that the (prospect of) engaging patients in their own care process seems to incentivize healthcare professionals to implement PROMs in clinical care, which subsequently may lead to a more individualized health service delivery and an improvement in the quality of care.

The presence of a PROM coordinator in the (outpatient)clinic, which some respondents felt should be the role of a nurse practitioner, was a frequently reported facilitator. Similarly, "the absence or lack of staff for coordination" was also a reported barrier in the current study. This factor may be related to other barriers that were frequently reported, namely "time constraints and/or the "labor-intensive nature of the implementation". Antunes et al.[8] and Dunckley et al.[33] also identified a coordinator as being essential for implementing PROMs in palliative care. In collaboration with the VBHC expertise team of the Erasmus University Medical Center, the coordinator assembles the members of the relevant disease team and explains what PROMs are and what is expected from them. In addition, the coordinator should be the linking pin between different levels in the organization (e.g. between management and in daily clinical practice) and support the multidisciplinary team. The coordinator also ensures that the multidisciplinary team comes together, prepares the office hours, and is the point of contact for the IT team that deals with the construction and implementation of the IT-system[8].

Issues related to the IT infrastructure were also frequently reported as a barrier, with respondents reporting difficulty in accessing the web-based PROM platform, the slow working speed of the platform, the visually unappealing dashboard, and the lack of integration in the EHR. Correspondingly, van Egdom et al.[10] found the

integration of PROMs in the EHR to be a facilitator for implementing PROMs in specifically breast cancer care. Electronic capture of PROMs in the clinic and between appointments allows real monitoring of symptoms, flexible scheduling of hospital appointments in response to PROM data, early detection of problems, and prompt clinical interventions.[34] However, the PRO data needs to be updated frequently and made accessible to patients and healthcare professionals. The use of PROMs in daily clinical practice also requires an adequate IT-infrastructure for the design and visualization of these PROs in a clear and insightful way, in order to provide meaningful feedback to the patient and running smoothly in the consultation room. In addition, it should be able to show a trend in PROM scores after treatment to give patients insight into their own course of the disease.[21,34]

The presence of a "language barrier for patients with a primary language other than Dutch " was the only barrier identified through the questionnaire statements. In a study by Schamber et al.[34], non-native patients showed a trend towards a lower rate of survey completion although not statistically significant. As one of the potential benefits of PROMs is the discussion of patient issues that otherwise would not have been identified, it is crucial in regards to healthcare equity that the needs of non-native speakers and patients with low health literacy skills are also taken into account. The availability of PROMs in multiple languages may be one effective and relatively inexpensive way of decreasing the reporting gap between native and non-native speakers.

Some overlap was observed between the responses to questionnaire statements and the open-ended questions. Overlapping facilitators included the confidence that implementing PROMs would lead to improved/more efficient patient care, the presence of a PROM coordinator, and a well-functioning IT infrastructure. The only overlapping barrier was the "language barrier for patients with a primary language other than Dutch". These aforementioned factors were the most prominent in the current study, and need to be especially considered prior to and during future implementation plans.

Strengths and Limitations

A key strength of this study was the prospective nature of the data collection. In addition, while previous research on the implementation of PROMs as summarized by Foster et al. has typically focused on a particular disease (e.g. cancer)[35], the field of healthcare (e.g. palliative care)[8] or a specific phase of the implementation process[9], the current study intentionally included all phases of the implementation process. Furthermore, the respondents were involved in the care of a variety

of diseases, albeit selectively chronic conditions. The web-based questionnaire's combination of both closed and open-ended questions was another strength of this study, as it gave respondents the opportunity to contextualize their responses to the questionnaire statements and report additional facilitators and barriers not included in the statements. Despite the combination of both types of questions, this cross-sectional study was unable to delve deeper into underlying mechanisms between facilitators or barriers and the implementation process. Future research with focus groups of healthcare professionals and administrative personnel may provide further insights into these mechanisms.

However, several limitations should also be mentioned. Firstly, the questionnaire used in this study was an adapted version of the "Barriers and Facilitators Assessment Instrument", a questionnaire previously developed and validated by Peters et al.[24]. This adapted questionnaire, which consisted of revised or newly added statements/ questions specific to the study's aim and setting, has not been validated or used in previous research. We acknowledge that, in retrospect, some statements (e.g. "The purpose of implementing PROMs is clear") could have been interpreted by respondents as a general statement with which they could agree or disagree, and not as a facilitator/ barrier that they have actually experienced as such in practice. Nonetheless, we still believe that these statements embody important factors that should be taken into account before and throughout the different stages of PROM implementation. Secondly, this study only involved healthcare professionals of a single academic medical center, which typically has more clinically complex patients, higher resource utilization, and a different workflow than e.g. a general hospital. This may limit the generalizability of the healthcare professionals' perspectives as identified in this study. Thirdly, only 61 respondents out of 187 invitees responded, despite attempts to increase the response rate by sending out reminders twice. On the other hand, the response rate across different specialties was high considering the fact that 17 specialties/ disease teams were represented by at least one respondent. Lastly, as it is conceivable that the respondents are relatively highly motivated employees who favor the use of PROMs for the potential quality of care improvement for their patients, the results may somewhat be biased. Specifically, it is possible that respondents were more likely to have a positive attitude towards PROMs implementation and view certain factors as facilitators, which may also explain why only one barrier was experienced from the questionnaire statements. The responses to the open-ended questions, however, did mention a considerable number of barriers as well.

CONCLUSIONS

In this study, commonly reported facilitators for implementing PROMs in routine clinical care were the presence of a coordinator, intrinsic motivation of members within a multidisciplinary disease team, and the integration of PROMs in the EHR. On the other hand, frequently reported barriers were time constraints, IT issues, and language barriers for patients with a primary language other than Dutch. In general, these findings are consistent with previous research. It is imperative that healthcare organizations that plan to incorporate PROMs in routine clinical care, account for these facilitators and barriers prior to and during the implementation process.

CONFLICT OF INTERESTS

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

CONTRIBUTORS

Authors MA, AO, NvL, LV and JH were involved in the conception and design of the study. The utilized questionnaire was constructed and reviewed by MA and NvL. MA and AO undertook the analyses and prepared the first draft of the manuscript. JH, HL, FE, LV and NK contributed to the interpretation of study findings and revised the manuscript. All authors critically reviewed and approved the final manuscript.

REFERENCES

1. Rivera SC, Kyte DG, Aiyegbusi OL, Slade AL, McMullan C, Calvert MJ. The impact of patient-reported outcome (PRO) data from clinical trials: a systematic review and critical analysis. *Health and quality of life outcomes.* 2019;17(1):156.
2. Porter ME. What is value in health care? *N Engl J Med.* 2010;363(26):2477-2481.
3. Porter ME, Larsson S, Lee TH. Standardizing Patient Outcomes Measurement. *The New England journal of medicine.* 2016;374(6):504-506.
4. Brédart A, Marrel A, Abetz-Webb L, Lasch K, Acquadro C. Interviewing to develop Patient-Reported Outcome (PRO) measures for clinical research: eliciting patients' experience. *Health and quality of life outcomes.* 2014;12:15-15.
5. Denniston AK, Kyte D, Calvert M, Burr JM. An introduction to patient-reported outcome measures in ophthalmic research. *Eye (Lond).* 2014;28(6):637-645.
6. Detmar SB. Use of HRQOL questionnaires to facilitate patient-physician communication. *Expert Rev Pharmacoecon Outcomes Res.* 2003;3(3):215-217.
7. Ruseckaite R, Maharaj AD, Krysinska K, Dean J, Ahern S. Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries. *Patient Relat Outcome Meas.* 2019;10:355-372.
8. Antunes B, Harding R, Higginson IJ, Euroimpact. Implementing patient-reported outcome measures in palliative care clinical practice: a systematic review of facilitators and barriers. *Palliative medicine.* 2014;28(2):158-175.
9. Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. *BMJ Qual Saf.* 2014;23(6):508-518.
10. van Egdom LSE, Oemrawsingh A, Verweij LM, et al. Implementing Patient-Reported Outcome Measures in Clinical Breast Cancer Care: A Systematic Review. *Value Health.* 2019;22(10):1197-1226.
11. Appleby J, Devlin N. *Measuring NHS Success: Can Patients' Views on Health Outcomes Help to Manage Performance?* : King's Fund; 2005.
12. Basch EM, Abernethy A, Mullins CD, Tiglao MR, Tunis SR. Ev1 development of a guidance for including patient-reported outcomes (Pros) in post-approval clinical trials of oncology drugs for comparative effectiveness research (Cer). *Value in Health.* 2011;14(3):A10.
13. Squitieri L, Bozic KJ, Pusic AL. The Role of Patient-Reported Outcome Measures in Value-Based Payment Reform. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2017;20(6):834-836.
14. Ong WL, Schouwenburg MG, van Bommel ACM, et al. A Standard Set of Value-Based Patient-Centered Outcomes for Breast Cancer: The International Consortium for Health Outcomes Measurement (ICHOM) Initiative. *JAMA oncology.* 2017;3(5):677-685.
15. van Egdom LSE, Lagendijk M, van der Kemp MH, et al. Implementation of Value Based Breast Cancer Care. *Eur J Surg Oncol.* 2019;45(7):1163-1170.
16. Dronkers EAC, Baatenburg de Jong RJ, van der Poel EF, Sewnaik A, Offerman MPJ. Keys to successful implementation of routine symptom monitoring in head and neck oncology with "Healthcare Monitor" and patients' perspectives of quality of care. *Head & neck.* 2020.
17. Arora J HM. Implementing ICHOM's Standard Sets of Outcomes: Cleft Lip and Palate at Erasmus University Medical Centre in the Netherlands. London, UK: International Consortium for Health Outcomes Measurement (ICHOM) (available at www.ichom.org). 2016.

18. Salinas J, Sprinkhuizen SM, Ackerson T, et al. An International Standard Set of Patient-Centered Outcome Measures After Stroke. *Stroke.* 2016;47(1):180-186.
19. Mulder J, Galema-Boers AMH, de Jong-Verweij LM, Hazelzet JA, Roeters van Lennep JE. The development and first results of a health-related outcomes set in familial hypercholesterolemia (FH) patients: Knowledge is health. *Atherosclerosis.* 2019;293:11-17.
20. van den Hoven AT, Bons LR, R HM, et al. A Value-based Healthcare approach: Health Related Quality of Life and Psychosocial Functioning in Women with Turner Syndrome. *Clin Endocrinol (Oxf).* 2020.
21. Foster A, Croot L, Brazier J, Harris J, O'Cathain A. The facilitators and barriers to implementing patient reported outcome measures in organisations delivering health related services: a systematic review of reviews. *J Patient Rep Outcomes.* 2018;2:46.
22. Garcia-Retamero R, Cokely ET. Communicating health risks with visual aids. *Current Directions in Psychological Science.* 2013;22(5):392-399.
23. Lugtenberg M, Burgers JS, Besters CF, Han D, Westert GP. Perceived barriers to guideline adherence: a survey among general practitioners. *BMC Fam Pract.* 2011;12:98.
24. Peters MAJ, Harmsen M, Laurant MGH, Wensing M. Ruimte voor verandering? Knelpunten en mogelijkheden voor verbeteringen in de patiëntenzorg. Uitgave: Afdeling Kwaliteit van zorg (WOK). *Nijmegen, the Netherlands: UMC St Radboud.* 2003.
25. Harmsen M, Peters M, Wensing M. Barriers and facilitators assessment instrument: Introduction, instructions and instrument. *IQ Healthcare.* 2005.
26. Fletcher J. Ethical approval for all studies involving human participants. *CMAJ.* 2015;187(2):91-91.
27. Santana MJ, Haverman L, Absolom K, et al. Training clinicians in how to use patient-reported outcome measures in routine clinical practice. *Quality of Life Research.* 2015;24(7):1707-1718.
28. Rubenstein LV, McCoy JM, Cope DW, et al. Improving patient quality of life with feedback to physicians about functional status. *J Gen Intern Med.* 1995;10(11):607-614.
29. Gutteling JJ, Darlington A-SE, Janssen HLA, Duivenvoorden HJ, Busschbach JJV, Robert A. Effectiveness of health-related quality-of-life measurement in clinical practice: a prospective, randomized controlled trial in patients with chronic liver disease and their physicians. *Quality of Life Research.* 2008;17(2):195-205.
30. Velikova G, Booth L, Smith AB, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. *Journal of Clinical Oncology.* 2004;22(4):714-724.
31. Søreide K, Søreide AH. Using patient-reported outcome measures for improved decision-making in patients with gastrointestinal cancer–the last clinical frontier in surgical oncology? *Frontiers in oncology.* 2013;3:157.
32. Cameron E, Green M. *Making sense of change management: A complete guide to the models, tools and techniques of organizational change.* Kogan Page Publishers; 2019.
33. Dunckley M, Aspinal F, Addington-Hall JM, Hughes R, Higginson IJ. A research study to identify facilitators and barriers to outcome measure implementation. *International journal of palliative nursing.* 2005;11(5):218-225.
34. Schamber EM, Takemoto SK, Chenok KE, Bozic KJ. Barriers to completion of Patient Reported Outcome Measures. *The Journal of arthroplasty.* 2013;28(9):1449-1453.
35. Howell D, Molloy S, Wilkinson K, et al. Patient-reported outcomes in routine cancer clinical practice: a scoping review of use, impact on health outcomes, and implementation factors. *Annals of oncology : official journal of the European Society for Medical Oncology.* 2015;26(9):1846-1858.

SUPPLEMENTAL MATERIAL

Supplementary Table 1 | Involved Disease Teams of Erasmus University Medical Center Rotterdam.

No.	Disease teams	Phases	No.	Disease teams	Phases
1	Esophagus and anorectal atresia	I	20	Obesity	P
2	Inflammatory bowel disease (IBD)	D	21	Pregnancy and birth care	P
3	Primary Immune Deficiencies	D	22	Cervical carcinoma	C
4	Craniofacial microsomia	D	23	Eczema	C
5	Cystic fibrosis	D	24	Functional bladder disorder	C
6	YEAST	D	25	Pediatric chest surgery	C
7	Liver failure	D	26	Liver tumors	C
8	Multiple sclerosis	D	27	Multiple myeloma	C
9	Staphylococcus aureus bacteremia (SAB)	D	28	Primary ovarian insufficiency (POI)	C
10	Obstruction icterus	D	29	Bladder cancer	R
11	Peripheral arterial disease	D	30	Breast cancer	R
12	Kidney donations / transplantation	D	31	Familial hypercholesterolemia	R
13	Sarcomas	D	32	Cerebral infarction (stroke)	R
14	Testicular sperm extraction (TESE)	D	33	Brain tumors	R
15	Bronchiectasis	P	34	Ear, nose & throat disorders (ENT)	R
16	Congenital hand abnormalities	P	35	Cleft lip and palate	R
17	Endometriosis	P	36	Turner syndrome	R
18	Lung cancer	P	37	Sickle cell children	R
19	Melanomas	P	38	Sickle cell adults	R

Note. I = “Purpose and Intake”; D = “Designing”; P = “Preparing”; C = “Commencing”; R = “Reflecting and Developing”

Supplementary Table 2 | Professional Responses on Different Open-Ended Questions Relevant to PROMs Implementation in Clinical Care. Exemplary Quotes Were Selected to Illustrate Recurrent Thematic Elements.

What did you experience as the biggest promoting factor for the implementation of PROMs in the (outpatient) clinic?
- Enthusiasm and motivation within team are the drive to set up something well. - Enthusiasm of patient to be involve in own care. - Team building around a clinical condition and joint thinking and communication about definition. - The collaboration with multiple disciplines gives positive energy. - The help of the quality advisor who makes the link between daily practice and IT. - The fact that all involved care providers fully support the introduction of the PROMS. - Patient's support by understanding the extent of benefits for them (better insight in patient psychosocial problems, better care, better monitoring). - Patient participation and appreciation. - Support from 'value-based healthcare' team in hospital. - Well-organized IT. - Experience gained from pioneers within the outpatient clinic. - Clear information/ knowledge, guideline, and way of interpretation of PROMs. - Providing enough resources like iPad at the clinic for employees.
What did you experience as the biggest impeding factor for the implementation of PROMs in the (outpatient) clinic?
- Platform is slow functioning to record data and to properly retrieve and display. It takes a lot of time; but new problems come to light every time and it feels every time the wheel is reinvented. - E-mail service does not work properly to send the questionnaires automatically. - There is no direct link between PROMs questionnaire and electronic health record. - The lack of a clear, intuitive, and visually appealing dashboard for physicians. - We receive not enough support from IT staff. - It is a lot of work behind the scenes, such as maintaining databases and sending questionnaires to patients - Takes a lot of time, especially sending emails, which is often not answered - Time has to be divided between patient care and administrative tasks with a lot of registration and logistic burden. - Preparing with PROMs takes a little more time. - There is insufficient time to make appropriate agreements with patient. - Lack of employees, who can complete the questionnaires with patient on iPad prior to visit hours. - Lack of support in actual start-up and implementation in terms of logistics. - Insufficient support of nurse practitioner/physician assistant. - The entire team has not the same motivation to work with PROMs. - Manager was not enthusiastic. - Cooperating parties have different interests. - There is not enough communication within team. - Learn new routines in the use of the questionnaires. - Diversity and lack of uniformity in implementation. - It was a very long, uncluttered process in which the rules were constantly changing throughout the implementation. - We do not have an email address for all patients prior to the first outpatient visit. This means that not everyone can participate in completing the questionnaires prior to the visit. - It is also not always possible for elderly patients and patients who do not have a good command of Dutch language to complete the PROMs. - Supporting groups is young. - Colleagues that work from an older paradigm.

Supplementary Table 2 | Professional Responses on Different Open-Ended Questions Relevant to PROMs Implementation in Clinical Care. Exemplary Quotes Were Selected to Illustrate Recurrent Thematic Elements. (*continued*)

Can you name one or more circumstances / factors that would make it easier for you to implement PROMs structurally?
- Improvement of IT structure to measure PROMs and access results quicker. - Link PROMs directly to the patient electronic health record. - A visually appealing dashboard to make enthusiasm in care provider and patient. - Link questionnaires directly to an outpatient appointment instead of sending them separately. - A faster platform to view dashboard results and quicker integration of results with clinical data. - A clear, well-thought combined platform that is appropriate for medical research and easy data extraction. - More supporting staff who can help to fill in PROMS prior to the consultation and keep track of everything from the start of treatment to the end. - Logistical assurance of the administrative process. - Assign extra consultation time to support in obtaining and directing data. - Ensure there is a desk that organizes the whole story from questionnaire to dashboard. - Permanent PROMs employees at all outpatient clinics who take care of the enormous administrative burden. - Presence of a nurse practitioner/physician assistant support. - If the administrative entire process runs automatically / digitally as much as possible. - Less 'clicks'; better overview. - Better link with appointments. It should be easier to have a clear picture of what you want to collect and for what purpose. - Be visible whether a questionnaire has been sent and completed and the automatic sending of a questionnaire by appointment code. Also, an automatic reminder if has not yet filled in, would also be nice. - Send PROMs directly from patient electronic health record and that is also clear to every individual patient where he / she is in the care path and on which measuring time point. - Having access to a clear and comprehensive e-learning for care providers. - Having a better insight into the use of the patient portal and ways of informing / contacting patients. - Having awareness on research results in this area which showing what the use of PROMs adds in practice. - Clear idea of what one wants to measure and why. - Motivating patients. - Explicit attention paid to vulnerable patients, language barriers; visual and auditive support. - Structured processes; build on experiences of other institutions. - Good communication and short consultation structures. - Providing sufficient finances / resources.

Supplementary Table 2 | Professional Responses on Different Open-Ended Questions Relevant to PROMs Implementation in Clinical Care. Exemplary Quotes Were Selected to Illustrate Recurrent Thematic Elements. (*continued*)

Can you name one or more circumstances / factors that would make it more difficult for you to implement PROMs structurally?
- Insufficient administrative staff to properly organize healthcare and monitor whether questionnaires are being filled out. - If the IT is incapable, no good IT infrastructure for dashboard and digital questionnaires. - No patients support: if they notice that nothing is being done with the PROMs, they will no longer complete the PROMs questionnaires. - If clinicians do not see the necessity or advantages of implementing PROMs and stop to use it. - Implementation of PROMs costs substantial money. Lack of financial support would make it difficult to implement. - Defensive or unmotivated colleagues. - If time duration for visiting and consultation with patient would not be enough. - If the IT lags behind. - If there is no direct link with patient's electronic health record. - Little motivation among the team members. - If there were no appropriate infrastructure and clearly designed care paths aimed at the PROMs outcomes. - Not enough knowledge which PROMs is most appropriate. - Not enough experience on how the implementation works. - Lack of collaboration with external parties. - More difficult if there's official consent needed from patients.
Based on your experience, what do you see as the biggest advantage of implementing PROMs in the (outpatient) clinic?
- Taking a complete picture of the patient situation before an outpatient visit. The PROMs highlight the elements that the patient finds important and recognize complaints which patient would normally not bring up. - PROMs are structured inventory important outcomes to determine how patients react to treatment - Taking a structured feedback from the patient about how he / she experiences illness and treatment. - Insight over added value for the patient. - Preparing patient and care provider for structured and data-driven conversation at outpatient visit and offers patient the opportunity to indicate problems / ask questions. - Better insight into patient need / wishes / goals, patient satisfaction, and treatment outcome. - More attention to patient' s quality of life, psychosocial, and social emotional. - Opportunity to provide a customized and appropriate care for patient. - Opportunity to provide more holistic and personal approach to the patient. - Better monitoring of any interventions effect. - Integration of PROMS in the discussion with patient could be used in shared decision-making and in setting up goals and prioritize care. - Visualization of PROMs would make the provided quality of care more efficient and it would be easier to get to the point. - Being able to provide care based on what the patient finds important or is inconvenienced by. - Speeding up reading of outcomes in the doctor's office and use for reporting to a medical specialist. - Providing a lot of data for research. - Researching trends in the larger group of patients. - Broader view of patient management using PROMS. - Evaluation of own performance, measuring change in outcome after improvements in treatment. - Enable multidisciplinary approach.

Supplementary Table 2 | Professional Responses on Different Open-Ended Questions Relevant to PROMs Implementation in Clinical Care. Exemplary Quotes Were Selected to Illustrate Recurrent Thematic Elements. (*continued*)

Based on your experience, what do you see as the biggest disadvantage of implementing PROMs in the (outpatient) clinic?
- Both implementation and using PROMs in outpatient clinic is very time-consuming. - Sending the questionnaires, interpretation, and providing feedback from completed questionnaires are very time-consuming. - It takes the time to review scores beforehand. - It takes time to introduce PROMs within team and develop it. - It takes the time to generate information for the dashboard. - Over-diagnosis is possible by asking various subjects, e.g. if psychosocial problems related to illness are questioned and patients indicate problems, care must also be taken to provide adequate help or referral (health services). - Registration burden for patients; difficult to motivate them - PROMs questionnaires are not optimal for some diseases like congenital disease. - Organization burden especially in busy clinics with insufficient support. - Misinterpretation of PROMs results is possible since physicians work differently. - Finance burden. - Higher workload. - A whole new work style will have to be taught. - Doctors are asked to adopt a new workstyle; if PROMs are wrongly interpreted or not implemented correctly, then the PROMs become worthless. - If questionnaires are filled out once a year, scores may become obsolete at consultation. - No flexibility in filling out of questionnaire and interpretation. - It is difficult to explain meaning of scores and there is no clarity about benchmarking certain questionnaires. - Unknown interest for the larger public.
What would be the main reason (s) for you to implement PROMs in the (outpatient) clinic?
- More participation of patient in care. - Improving efficiency of outpatient appointment and having structured consultation with patient. - Having a better insight into patient wishes, needs, and experiences and aligning care with patient wishes. - Stimulate patients to health-promoting behavior. - A more personalized care for patient and more structured consultation hours. - Having a better insight into patient quality of life. - Providing an appropriate tool for comparison of patient heath status. - Involving the patient in the care process and being able to better respond to the actual demand of the patient. - Adjusting outcomes if necessary, for a better quality of life / experiences of the patient. - Early and quick recognition of complaints / symptoms. So can respond to it on time and possibly refer or give advice on what to do with complaints. - PROMs measurement is handy tool for treatment decision-making and set goals. - PROMs measurement is holistic approach and provides room for non-medical aspects. - Uniformity in approach which can lead to a patient-tailored approach. - PROMs are patient friendly and promote patient input. - Bringing together specific condition (e.g. obstetric) and psychiatric information. - Evaluation of own functions and possibility to use the information during the consultation. - Ability to easily list outcomes and have insight in the entire patient population. - Support for research, scientific research is possible by evaluating structured collection of data over time. - More labor satisfaction.

Supplementary Table 2 | Professional Responses on Different Open-Ended Questions Relevant to PROMs Implementation in Clinical Care. Exemplary Quotes Were Selected to Illustrate Recurrent Thematic Elements. (*continued*)

What would be the main reason (s) for you not to implement PROMs in the (outpatient) clinic?
- The amount of extra time it takes to read, interpret, and register everything afterwards. - Time investment for maintenance of the platform. - If it delays our workflow. - Time investment for extra administrative tasks. - Insufficient support in the formulation, adaptation, and sending out questionnaires. - Insufficient personnel, financial, or organizational support. Too little motivation to organize care differently. - No patient supports because of language barrier or if patient would deliberately fill out worse scores for certain treatments. - Insufficient IT support and rigid infrastructure. - Lack of team or organization support. - The platform is not user-friendly. - The questionnaires are lengthy for patients and too much burden for patients. - Increasing administrative burden and may not added value for the patient. - If we cannot respond patient demands appropriately. - If extra work / costs outweigh the benefits. - If it will impede the workflow.

Chapter 9 General Discussion

The overall aim of this dissertation has been two-fold: to study methodological aspects of the analysis of patient-reported outcome measures (PROMs), and to explore the implementation and application of PROMs in clinical care. PROMs can be defined as *instruments that measure aspects of a patient's health condition and/ or treatment that come directly from the patient, thus without external interpretation.*[1] These aspects of (health-related) quality of life can be quantified for comparative effectiveness research, assessing the quality of care, and benchmarking.[2]

This chapter includes a synthesis (Box 9.1) and reflection of the key findings of included studies, which will be discussed separately for the methodological aspects of PROMs (Part I) and the practical applications of PROMs in clinical care (Part II). In the last paragraph, perspectives on future research and applications of PROMs in clinical care are described. An overview of key recommendations (Box 9.2), distilled from the studies in this dissertation, conclude this chapter.

Box 9.1 | Overview of Key Findings Per Research Question

Questions	Answers
Part I: Methodological Aspects of PROMs	
Which general patient-reported morbidity instruments have been published (in order to use for case-mix adjustment)?	A systematic review was conducted and ultimately included 34 articles. The Self-Reported Charlson Comorbidity Index, Self-Administered Comorbidity Questionnaire and Disease Burden Morbidity Assessment were the most frequently used instruments. Commonly included conditions were diabetes, hypertension and myocardial infarction. Studies demonstrated substantial variability in item-level reliability versus medical record review, meaning that the accuracy of patient-reported morbidity data is dependent on the specific disease/condition in question.
Which predictors are important for case-mix adjustment models for a generic patient-reported outcome measure (EQ-5D) within acute ischemic stroke care, and how does the model differ from models for clinical outcomes (mortality, modified Rankin Scale)?	Case-mix models were developed using a data subset of the Dutch National Stroke Registry. In contrast to the models for clinical outcomes, more statistically significant predictors were identified for the case-mix model for EQ-5D with exclusive predictors being nationality, sex, and socio-economic status. The predictive ability of the case-mix model for the PROM was slightly less than the models for mortality and the modified Rankin Scale.
What are the differences in case-mix adjustment models between individual domains of a generic patient-reported outcome measure (EQ-5D) within acute ischemic stroke, and what is the influence of case-mix adjustment on the reliability of individual EQ-5D domains as quality indicators for between-hospital comparisons in the quality of ischemic stroke care?	Case-mix models for individual EQ-5D domains were developed using selected MR CLEAN trial data. A clear difference between the "physical" domains (self-care, mobility, usual activities) and "psychological" domains (pain/discomfort, anxiety/depression) was observed with less significant predictors being identified for the latter domains. The predictive ability of the case-mix models for psychological domains were also substantially less than those for physical domains. After case-mix adjustment, hospital effect size estimates only slightly changed for physical domains as opposed to psychological domains.

Box 9.1 | Overview of Key Findings Per Research Question (continued)

Questions	Answers
Part II: Practical Applications of PROMs in Clinical Care	
How have patient-reported outcome measures been administered in clinical breast cancer care, what is its impact on patients, care providers and care processes, and what are the reported facilitators and barriers?	From 34 articles included in a systematic review, a variety of methods (mostly electronically) and frequencies of PROM administration/ collection was observed across multiple settings (from screening to treatment surveillance). Most studies reported that systematic PROM collection had a promising impact on patients, care and care process/ system outcomes. Some studies also described experienced facilitators and barriers of the collection process, which mostly revolved around technical (e.g. IT-infrastructure), behavioral (e.g. change champion) and other (e.g. time constraints in the clinic workflow) factors.
What is the course of patient-reported outcome measurement scores over time within (surgical) treatment strategies for breast cancer?	Prospectively collected PROMs for breast cancer patients at baseline (prior to surgery), T6 (6 months post-primary surgery) and T12 (12 months post-primary surgery) were analyzed. From baseline to 12 months follow-up, there was a significant improvement in "Body Image" (EORTC-QLQ-BR23) for patients following BCT, while "Physical Wellbeing" (BREAST-Q), "Physical Functioning" (EORTC-QLQ-C30), "Sexual Wellbeing" (BREAST-Q) all significantly declined compared to baseline. Median "Satisfaction with Breasts" (BREAST-Q) and "Psychosocial Wellbeing" (BREAST-Q) scores also significantly declined at T12 for patients who underwent mastectomy without reconstruction.
What are the normative BREAST-Q scores in a Dutch population sample?	A cross-sectional study was conducted by inviting Dutch women to complete a web-based translated version of the BREAST-Q instrument. 9059 women without breast cancer in prior history were surveyed. Mean BREAST-Q domain scores were described for "Satisfaction with Breasts", "Psychosocial Wellbeing", "Physical Wellbeing" and "Sexual Wellbeing". Domain scores notably followed similar patterns to internationally published US normative data. Prior non-cancer-related breast surgery was a positive predictor for "Satisfaction with Breasts" and "Sexual Wellbeing".
What are the facilitators and barriers of implementing patient-reported outcome measures in clinical care, as reported by healthcare professionals of an academic center?	Healthcare professionals from one academic medical center were invited to complete a customized online questionnaire on their perspectives on the facilitators and barriers of the implementation of PROMs in clinical care. 61 respondents (response rate 33%) from both surgical and non-surgical departments were included. Frequently mentioned facilitators were the presence of a PROM coordinator in the clinic, the multidisciplinary collaboration, and the integration of PROMs in the electronic health record. Barriers that were commonly reported were IT-issues, and time constraints/ labor intensive nature of the collection and review process.

PART I – METHODOLOGICAL ASPECTS OF PROMS

Part I of this dissertation has not only focused on patient-reported multi-morbidity instruments as a potential data source for this important case-mix factor, but also on case-mix adjustment models for aggregated PROMs.

In addition to traditional outcomes for healthcare quality assessment (e.g. mortality, complications, rehospitalizations), there has been an increase in the collection

of patient-reported outcome measures (PROMs). There are still substantial knowledge gaps in the methods (how?) and the appropriate settings (when?) for proper case-mix adjustment of PROMs.

Patient-Reported Morbidity

The rationale for the systematic review in **Chapter 2** on patient-reported morbidity instruments, used for capturing morbidity as a case-mix variable, related firstly to the lack of consensus among clinicians on the best source of morbidity data[3] (e.g. patient self-report[4], health record review[5], diagnosis codes[6], claims data[7], pharmacy data[8]), and secondly to provide an overview of (validated) instruments to assist interested clinicians in selecting a measure suitable for their clinical setting. Measures of morbidity are frequently used within health services research to control for confounding, to identify effect modifiers, to predict outcome and to improve statistical efficiency.[9] Morbidity indices are members of the broad family of case-mix adjustment systems, which have traditionally been developed for risk adjustment of outcomes such as mortality, adverse events or health care utilization (or different combinations of them).

During the development of the literature search strategy, it became apparent that certain terminology was used interchangeably across the screened abstracts. Concepts and definitions of "comorbidity", "morbidity", and "multimorbidity" are often not consistently used in the literature.[10] "Comorbidity" is currently understood in relation to an index disease.[11] This definition automatically sparks the debate on which disease or condition should be designated the index disease or condition and which the comorbid disease or condition. The choice of an index disease has varied in relation to a specific research question, the disease that elicited initial care, or the attending physician's specialty. "Morbidity" is regarded as the presence of diseases or conditions. There have been various conceptualizations for "multimorbidity" throughout the literature: the number of co-occurring diseases or medical conditions, the number and severity of diseases or conditions, and the number and severity of diseases or conditions with their net impact on functional status.[12] As the number of patients with multi-morbidity will increase dramatically due to aging of the population in the Netherlands, it will be essential to move away from the single disease focus that has prevailed in clinical medicine for centuries. Therefore, it will also be required for both medical research and clinical practice to consistently use appropriate nomenclature. Considering the multitude of concepts and definitions, the literature search strategy was jointly created with an experienced librarian in order to optimize the sensitivity of the

search by combining thesaurus terms (Emtree terms for Embase and MeSH terms for Medline) with relevant terms in the titles and abstracts.

In the systematic review, various patient-reported morbidity instruments were identified, of which the Self-Reported Charlson Comorbidity Index (1996)[13], the Self-Administered Comorbidity Questionnaire (2003)[4], and the Disease Burden Morbidity Assessment (2005)[14] were the most frequently assessed and cited. Some included articles were validation studies of the aforementioned instruments. Most instruments formulated the question regarding presence of conditions/ diseases as "Do you have or have you ever had…?", "Has a doctor ever told you that you have…?", or "Do have you any of the following problems?", all with a close-ended response option. The length of instruments ranged from 4 items to 195 items. The response rate was described in nine studies, and ranged from 28% to 99%. Frequently included conditions/ diseases ("items") across the instruments were diabetes, hypertension, myocardial infarction, and stroke. The item-level reliability with other morbidity data sources was highest for diabetes and lowest for arthritis and stomach ulcers. These differences also accounted for the substantially wide κ value range between items across most studies. These observations can be explained by a variety of reasons: the condition/ disease not being diagnosed, the clinician not adequately disclosing the diagnosis to the patient, the patient forgetting (recall bias) or misunderstanding, or the patient being unwilling to report it. It's also possible that having multiple conditions/ diseases itself increases the likelihood of discordance. Whole instrument concordance with medical records was the highest, in contrast to medical-record derived morbidity indices and administrative data. Associations of patient-reported morbidity instruments' scores with clinical outcomes, PROMs and health care utilization outcomes were weak to fair, as demonstrated by a few studies.

Case-Mix Adjustment for PROMs

Only when patient characteristics like morbidity are collected, can potential case-mix adjustment be conducted for the assessment of healthcare quality.[15] Case-mix can be defined as patient characteristics that impact outcome but are beyond the control of the healthcare provider, such as age, gender, morbidity, severity of disease. Case-mix adjustment or correction is a statistical process that accounts for differences in patient characteristics across treatments, physicians or hospitals, in order to enable fair comparisons of outcomes of specific treatments or overall care provided.[16,17] Risk-adjustment is a crucial step for quality improvement, because otherwise differences in outcome may represent differences in baseline risk rather than quality of care when benchmarking hospitals.[18] Although many case-

mix adjustment models have been constructed and validated for clinical outcomes across multiple conditions[19,20], there still remains a scarcity of experience on the development of case-mix adjustment models for patient-reported outcome measures.[21,22] Recently, a strong desire has developed to include PROMs as outcomes for benchmarking.[23] The studies in **Chapters 3 and 4** highlight an important issue to consider before benchmarking PROMs namely the selection of appropriate case-mix factors, as these may differ than for traditional (clinical) outcomes.

In **Chapter 3**, data from ischemic stroke patients in the Dutch National Stroke Registry was used to develop a case-mix adjustment model for a generic PROM, the EQ-5D, and to compare it to case-mix adjustment models for a clinical (mortality) and a stroke-specific functional outcome (modified Rankin Scale; mRS). After backward selection, exclusive independent variables that were identified for the EQ-5D case-mix adjustment model were nationality (native Dutch vs. foreigner), sex, and socio-economic status (derived from status scores based on zip codes). Case-mix model performance for the EQ-5D was considerably less compared to the models for mortality and mRS. In contrast to the models for mortality and mRS, more predictors were required in the model for the EQ-5D for the predictive ability to reach a plateau. This observation might be explained by the lack of other baseline predictors such as baseline psychological, social and quality of life factors, which may contribute more to model performances for subjective ("abstract") outcomes like quality-of-life. The observation that model performances were better for outcomes that are somewhat more easily/ clearly interpretable ("concrete"), was also seen for when case-mix adjustment models were constructed for individual EQ-5D domains as outcomes. These observations reflect the idea that PROMs are more subjective and, therefore, influenced by a broader variety of factors than traditional (clinical) outcomes.

Chapter 4 describes a secondary analysis of ischemic stroke patients from the MR CLEAN trial[24] data, which included 12 stroke centers. The predictive ability of case-mix adjustment models for the psychological domains "anxiety/ depression" and "pain/ discomfort" was found to be substantially lower than for the physical domains of "usual activities", "mobility" and "self-care". More predictors were identified from univariable regression models for "self-care" (12 predictors), "mobility" (11 predictors) and "usual activities" (9 predictors) than for "pain/discomfort" (6 predictors) and "anxiety/depression" (4 predictors), giving the impression that different case-mix variables need to be included for more "abstract" domains in terms of interpretation. In addition, more stroke centers had a bigger change (≥ 50%) in effect sizes after case-mix adjustment for the psychological domains than for the physical domains. As most included hospitals had minimal changes in their effect estimates after case-mix adjustment, this substudy also suggests that differ-

ent case-mix variables (e.g. psychological and social factors) may be associated with PROMs than those traditionally collected for clinical outcomes.

The studies in **Chapter 3 and Chapter 4** illustrate the necessity of collecting non-traditional case-mix factors if the desire is to utilize PROMs as outcomes for benchmarking. These studies also suggest that some PROMs may be too "coarse" a measure (at certain time points) to actually differentiate between treatments or care providers within ischemic stroke care, especially when traditional case-mix factors have a minor influence on them.

PART II – PRACTICAL APPLICATIONS OF PROMS IN CLINICAL CARE

Patient-reported outcome measures (PROMs) were originally developed for clinical trials as secondary outcomes to weigh against traditional outcomes (e.g. survival), and for population studies as alternative expressions of health need.[25] Although PROMs were used to support treatment recommendations or inform health policy, there was not a direct benefit for those individual patients providing this data.[26] However, the clinical value of these holistic and comprehensive assessments of treatment benefit on an individual patient level has become a topic of interest due to the growing experience in using these measures for identifying or monitoring symptoms, determining (health-related) quality of life, evaluating treatment outcomes, and strengthening shared decision-making in the consultation room.[26,27]

Part II of this dissertation has focused on the components and impact of PROM administration (including methods, frequencies, questionnaires used, and goals) in breast cancer care, the analysis and interpretation of longitudinally collected PROMs of breast cancer patients, the collection and analysis of normative BREAST-Q data, and the identification of facilitators and barriers of implementing PROMs in clinical care as perceived by healthcare professionals.

Use of PROMs in Clinical Breast Cancer Care

In a systematic review, presented in **Chapter 5**, an overview was provided on how PROMs have been collected within clinical breast cancer care including various methods and frequencies of PROM administration. A total of 2311 articles were screened, of which 34 were included in the final selection. These included studies were mostly prospective or retrospective cohort studies (N=15), followed by randomized controlled trials (N=10), qualitative studies (N=5) and cross-sectional stud-

ies (N=4). The most frequently used instruments across articles were the Common Terminology Criteria for Adverse Events (CTCAE), the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire module QLQ-C30 (EORTC-QLQ-C30) and the EQ-5D. In the majority of studies, the administration of PROMs was during treatment with PROMs being mostly collected electronically through web-based assessments or tablets. Only four studies had the electronic PROM system being integrated in the electronic health record. More than half of the included studies described summary reports of PROM data being sent to care providers prior to the patient consultation.

This systematic review demonstrated the feasibility of routine PROM collection in breast cancer care, and its promising impact on patients, care providers, and care processes. Firstly, a number of studies demonstrated patients' positive attitude towards PROM collection with a high acceptability of electronic administration methods and the length and number of questionnaires administered. Moreover, some studies also found that patients considered PROMs as helpful in discussing health problems with their care providers during the consultation. Secondly, a handful of studies showed that care providers found PROMs helpful in identifying patients' symptoms and other areas of need. Finally, both patients and care providers were generally positive about the impact of PROM collection on the patient-provider communication. In addition to the feasibility and impact, some included studies also described the facilitators and barriers of the PROM collection. Key facilitators were electronic methods for capturing PROMs, graphical reports of PRO data, consistent scaling of PROM scores for interpretation, the presence of a "change champion" in the clinic, and protocols for PROM administration. On the flipside, common barriers were technical issues including the integration of PROMs in the EHR, time constraints in the clinic workflow, and unclear questions/items within PROMs including stating the (symptom) recall period.

The growing use of electronic systems for PROM data collection and storage has led to this rapid increase in the body of literature on the utilization of PROMs in clinical cancer care.[28] Most of these studies, however, have used PROMs as outcomes for interventions other than the actual incorporation of PROMs in clinical care. In order to include studies in which the PROM collection itself was the intervention, the literature search strategy was also optimized with an experienced librarian. After comparing the different patient populations across the included studies, it became apparent that there was limited diversity in ethnicity, socio-economic status and educational background. This striking observation emphasizes the point that PROMs need to be designed and made more accessible to those vulnerable

patients with less health literacy and a higher probability of having limited access to healthcare. As PROMs may improve patient engagement with their own care process[29], these minority populations need to specifically targeted as the treatment compliance and health benefits could potentially be much greater. Because most of the included studies in this review not only dealt with breast cancer patients but other cancer types (e.g. lung and genitourinary cancer) as well, the facilitators and barriers reported by some included studies are most likely common across most cancer treatment trajectories which have incorporated PROMs.

Chapter 6 described an observational cohort study which analyzed three prospectively collected PROMs (BREAST-Q, EORTC-QLQ-BR23, EORTC-QLQ-C30) of a breast cancer cohort that underwent either breast-conserving therapy (BCT) (N = 143) or mastectomy without reconstruction (N = 70) at the Erasmus University Medical Center between October 2015 and July 2019. With the exception of significantly more gene mutation carriers undergoing mastectomy, baseline characteristics did not differ between surgical strategies. In patients who underwent BCT, a significant improvement in median "Body Image" score was observed at 12 months follow-up compared to baseline, while a significant deterioration was seen for "Physical Wellbeing", "Physical Functioning" and "Sexual Wellbeing". At 12 months following mastectomy without reconstruction, significantly reduced "Satisfaction with Breasts" and "Psychosocial Wellbeing" scores were reported compared to baseline. While a moderately strong negative correlation was observed at T12 between "Physical Wellbeing" and "Breast Symptoms" (EORTC-QLQ-BR23) for patients following BCT, "Satisfaction with Breasts" and "Body Image" were moderately correlated at T12 for those who underwent mastectomy. Across all three PROMs, respondents and non-respondents did not significantly differ in baseline patient characteristics.

The routine collection of these PROMs is a component of a value-based breast cancer care initiative in the Erasmus MC, which was designed in 2014 and initiated by a multidisciplinary breast cancer team in October 2015.[30] This multidisciplinary team (including surgical oncologists, medical oncologists, radiation oncologist, radiologists, plastic & reconstructive surgeons, pathologists, breast cancer nurse specialists, clinical geneticists, psychologists) defined an outcomes set – consisting of both provider- and patient-reported outcomes – based on the Breast Cancer Standard Set of the International Consortium for Health Outcome Measurements (ICHOM).[31] In addition, an electronic data collection tool (the "Zorgmonitor"[32] based on GemsTracker[33]) was developed for automatic distribution of PROMs at determined time points during the entire breast cancer care cycle. Furthermore,

this tool was integrated into the electronic health record in order to review PROM data of individual patients and to discuss them in the consultation room at the (outpatient) clinic. This data collection tool may be one of the key reasons why the completion rate of PROMs is considered quite well in this academic center. For example, the BREAST-Q response rate declined from 71% at T0 to 67% at T12 in the aforementioned breast cancer cohort. This "technical facilitator" was also mentioned in an included study of the systematic review in Chapter 5.

The cohort study in Chapter 6 deliberately only described the trend of short-term post-operative PROM scores *within* breast cancer surgical strategies instead of PROM scores *between* strategies. As noted before in Chapter 5, the trend of PROM scores cannot be easily attributed to one component of an intervention (treatment), as most current breast cancer treatments encompass multiple modalities (e.g. BCT plus standard irradiation). Due to the non-randomized nature of the data, it is very likely that outcome (including PROMs) differences between treatment strategies are confounded by known and unknown baseline demographic characteristics (e.g. age, socio-economic status, educational background), clinical characteristics (e.g. medical history, tumor grade, (neo-adjuvant) radiation and/ or systemic treatment), and the treatment indication itself ("confounding by indication")[34]. Lack of blinding and poor control for (un)measured confounding hamper the direct comparison of the surgical cohorts and, thus, establishing a causal relationship between surgical treatment and the trajectory of PROM scores.[35] However, there are techniques, like case-mix adjustment, that exist to eliminate these biases that hinder the interpretation of observational studies. The reason why case-mix adjustment was not attempted in the cohort study was the poor registration of potentially important case-mix factors in the database. Nonetheless, the analysis of prospectively collected PROMs *within* treatment strategies may provide insights in the course of important quality-of-life domain scores for both clinicians and patients (prior to starting treatment), which may be beneficial in facilitating the patient-provider communication and managing future patients' expectations. And although RCTs unquestionably hold several advantages over observational studies, most importantly the causal effect interpretation, it should be recognized that this longitudinal "real world" data usually has a broader generalizability than RCTS due to less restricted inclusion criteria and its derivation from multiple sources (e.g. EHR and patient surveys).[36]

Normative PROM Scores

The cross-sectional study in **Chapter 7** recruited Dutch women without a prior history of breast cancer (N = 9059) to complete the BREAST-Q, a surgery-specific

PROM, with the aim to ultimately compare this normative data to the scores of current (Chapter 6) and future breast cancer patients. Mean BREAST-Q domain scores were 64 ± SD 19 ("Satisfaction with Breasts"), 72 ± SD 16 ("Psychosocial Wellbeing"), 90 ± SD 12 ("Physical Wellbeing") and 60 ± SD 15 ("Sexual Wellbeing"). These domain scores resemble similar patterns described in previously published international norms.[37,38] Univariable regression analyses revealed age as a linear term to be associated with "Satisfaction with Breasts" and "Psychosocial Wellbeing". "Physical Wellbeing and "Sexual Wellbeing" were found to be a quadratic function of age, with models identifying inflection points of 34 years and 41 years at which "Physical Wellbeing" and "Sexual Wellbeing" values began to increase and decrease, respectively. Multivariable regression analyses demonstrated a positive association between prior non-cancer-related breast surgery and "Satisfaction with Breasts" and "Sexual Wellbeing", while a negative association was observed with "Physical Wellbeing".

This study reported weak significant associations between age and BREAST-Q domains, which indicate small differences (on the BREAST-Q scale of 0 – 100) between women of different age groups. Not only the statistical significance of PROM score differences between ages should be taken into account, but also the extent to which these differences are clinically meaningful and useful for patient management. A limitation of this study was that age and prior breast surgery were the only demographic variables collected. In retrospect, the collection of other potentially relevant characteristics – such as cup size, BMI, morbidities, SES, educational background, marital status – may have identified strong associations with these quality-of-life domains. On the other hand, such questions were intentionally not put in the web-based survey as they could have come off as too confrontational to respondents thereby possibly limiting the response rate. More efforts in clinically contextualizing the scores of breast cancer patients, either with MCIDs (minimal clinically important differences)[39] or comparisons to reference or normative scores[40], are needed. In doing so, both patients and care providers can jointly establish realistic treatment goals by taking patient preferences into account. Furthermore, the availability of MCIDs or normative scores may increase the uptake of PROMs by care providers in their clinical practice, as it has been shown that healthcare professionals value PROMs if they have an influence on their clinical decision-making process.[41]

Facilitators and Barriers of Implementing PROMs

In **Chapter 8**, a cross-sectional study was presented which aimed to identify the facilitators and barriers of the implementation of PROMs in clinical care as reported

by healthcare professionals from one academic medical center. Medical specialists, nurse(s) (specialists), paramedics and researchers, from multiple surgical and non-surgical medical departments involved in a Value-Based Healthcare initiative, were invited to complete a customized web-based questionnaire. This survey contained both questionnaire statements and open-ended questions in reference to various components of all stages of PROM implementation ("Purpose and Intake", "Designing", "Preparing", "Commencing", and "Reflecting and Developing").[42] The majority of 61 study participants indicated that the implementation of PROMs in clinical care was facilitated by the presence of a PROM coordinator in the clinic, the integration of PROMs in the electronic health record, and the intrinsic motivation of involved members. On the other hand, frequently reported barriers were language barriers with patients, IT-issues and time constraints. These findings are, to a large extent, consistent with findings from previous published studies.[43] One of the limitations was the relatively low response rate (33%), which may ironically be, among others, a reflection on another important barrier reported in other studies: healthcare professionals' level of knowledge and confidence about using PROMs.[44]

In contrast to previous studies that focused on a particular area of healthcare (e.g. palliative care)[45] or a specific aspect of the PROM implementation process (e.g. feedback of PROM scores in the consultation room)[41], this cross-sectional analysis provided the views of various healthcare professionals representing a variety of (non-)surgical conditions/ diseases. This study was borne out of a real-information void on the experiences of (para-)medical professionals who have either successfully or unsuccessfully attempted to implement routine PRO measurements in clinical care. As the Erasmus University Medical Center had the most and longest experience in routinely collecting PROMs in the region as part of its Value-Based Healthcare initiative, one of the initial goals of this study is to use the generated data from this academic center to inform surrounding community-based hospitals in the Rotterdam region who are planning or have recently started capturing PROMs.

GENERALIZABILITY OF STUDY FINDINGS

The studies in Chapter 3 and 4 on the case-mix adjustment of patient-reported outcome measures (PROMs) solely focused on a generic PROM, the EQ-5D, which was collected from ischemic stroke patients three months after either hospital discharge or trial randomization. These studies demonstrated that non-traditional

socio-demographic characteristics are associated with the EQ-5D after ischemic stroke. Generic PROMs such as the EQ-5D can differ across acute and chronic diseases in terms of, for example, responsiveness (sensitivity to change) and may thus not be suitable for all diseases/ conditions. However, the concept that different case-mix factors (such as socio-economic status, nationality, psychological and social factors) are significantly associated to quality-of-life domains, in contrast to clinical outcomes after acute diseases (e.g. ischemic stroke), can be generalized with caution to not only disease-specific PROMs (e.g. Neuro-QOL) but chronic conditions (e.g. breast cancer) as well.

The cross-sectional study in Chapter 8 on the facilitators and barriers of PROM implementation was unique in its inclusion of healthcare professionals who either managed to successfully implement PROMs in daily practice or failed to fully complete the implementation process. Therefore, the responses generated from the survey represented positive and negative experiences, adding to the transparency and trustworthiness of the findings. Study participants were also from both non-surgical and surgical departments and treated a variety of conditions/ diseases, which increased the generalizability of the results. Conversely, the generalizability was tempered by only surveying healthcare professionals from one academic medical center. As community-based and academic hospitals differ in complexity of patient populations, resource utilization and clinical workflow, this may also come across in the facilitators and barriers perceived or experienced by healthcare professionals attempting to incorporate routine PRO measurement in the (outpatient) clinic.

PERSPECTIVES ON FUTURE RESEARCH ON PROMS

Despite the ubiquitous use of PROMs in clinical care and research, there still remains a scarcity of peer-reviewed scientific papers on actual quality of care improvements. This paucity of evidence may be due to a lack of published clinical trials and longitudinal studies, caveats related to the analysis and interpretation of PROM scores, and challenges in the implementation of PROMs (e.g. technical in terms of data storage, behavioral in terms of compliance). This section contains suggestions for future research on patient-reported morbidity instruments, case-mix adjustment models for PROMs, and methods for handling longitudinal PROM data.

Patient-Reported Morbidity

To assess the impact of morbidity (the co-occurrence of multiple diseases/ conditions) on either clinical or patient-reported outcomes, it is necessary to measure it. Diagnoses of these diseases are recorded in different settings within healthcare: in primary care by the general practitioner, in secondary care by the medical specials during outpatient visits or hospital admissions, and when determining the cause of death. In addition, it is also possible for a diagnosis to be derived from pharmaceutical prescription records for a limited number of diseases/ conditions. However, it has been demonstrated that there can be considerable discrepancy between these various data sources. This is likely related to different definitions between various levels of healthcare. For instance, while hospital registries capture acute "events", general practice registries record "episodes" of illness that could also be related to events that happened a year (or more) prior. Conversely, acute events in the hospital registries may not be found in the GP registries within the year in question due to a registration delay at the GP or insufficient timely reporting from secondary care. The use of patient-reported morbidity instruments may circumvent these issues by enabling patients to become a reliable source of disclosing diseases or conditions that they are currently dealing with.

With the growing health literacy in the Netherlands[46], patients are becoming increasingly more aware of their own health status. This shift in awareness warrants future development of morbidity instruments as a potential alternative morbidity data source, especially given the lack of a single gold standard measure of morbidity.[3] Important considerations should be the purpose for which the instrument is developed, the underlying construct, whether it is based on individual conditions/ diseases or organ systems, and the data required for its estimation. Further research, in terms of prognostic modelling and validation studies, is also needed to determine whether patient-reported morbidity instruments can serve as a risk adjustment tool for observational outcomes data. If the goal is to use morbidity instruments to measure this important case-mix factor for the comparison of treatment strategies or hospitals, it is necessary that the morbidity instrument is validated for the population and outcome of interest (clinical, patient-reported or healthcare utilization). To reduce the level of intra-instrument variability regarding the item-level reliability, future morbidity instruments should not include conditions/ diseases that have been "resolved" in the past, those without symptoms, those with overlapping names (arthritis vs. osteoarthritis vs. rheumatoid arthritis) and ambiguous disease categories (e.g. coronary heart disease). It has been demonstrated that diseases with clear definitions (e.g. diabetes) and those that require ongoing treatment had a higher concordance with other morbidity data sources.[47-49]

Case-Mix Adjustment of PROMs

Although case-mix adjustment for clinical outcomes (e.g. mortality) are well established, there is still a substantial knowledge gap on an optimal methodology and its application for PROMs. This knowledge gap needs to be addressed as aggregated PROM data have been routinely collected by some countries as healthcare quality indicators.[50] PROMs can not only enable healthcare providers to assess and improve the performance within their own organization (internal benchmarking), but also to compare their performance with other healthcare providers (external benchmarking) in an effort to improve transparency. In doing so, more focus will be put on driving forward quality of care improvements of health services. Ideological differences and practical uncertainties about when and how case-mix adjustment should be applied have led to considerable variability in its use.[51] This ongoing debate has mostly focused on the adjustment of traditional healthcare indicators, not on PROMs. Globally, there has been an growing demand from different entities, like the Organization for Economic Co-operation and Development (*OECD*), for broader PROM use in not only clinical care but also in performance measurement and the (external) benchmarking of quality of care on hospital-level.[52]

The extent of case-mix adjustment for (aggregated) PROMs, like for any other outcome, should be dependent on the goal of the data analysis (internal vs. external benchmarking). For internal benchmarking, it may make more sense to take more care in the selection of case-mix variables to adjust for in the models. For example, if patients' socio-economic status (SES) contributes to the variation in PROM scores between health services within one hospital, and SES was adjusted for when examining (treatment) outcomes, then any variation due to the differing SES (with its known association to access to healthcare) between the services would be adjusted out rather than being used to help explain possible health inequity and inform quality improvement initiatives. On the other hand, if the variation in PROM scores between hospitals is related to differences in SES between the hospitals' patient population, not adjusting for SES may lead to hospitals with a considerable proportion of underprivileged patients (lower SES) to be potentially disadvantaged in comparison to hospitals that mainly serve patient populations with higher SES. Furthermore, as earlier hypothesized, there has also been a growing body of evidence which suggest that psychological and social factors can be underlying determinants influencing PROMs as well.[53] The implications of improper case-mix adjustment, especially for external benchmarking, can have major ramifications to those care providers disadvantaged by incorrectly adjusted healthcare quality indicators: diminished income from health insurance

companies (under pay-for-performance) and an inability to attract and retain both healthcare professionals and patients.[51] Considering these potential consequences, more empirical research is required to identify case-mix factors that are strongly associated with commonly used PROMs and to assess the differences of case-mix adjustment models between clinical outcomes and PROMs.

Another issue that needs to be tackled in future research is whether or not baseline PROM scores need to be added as predictors to case-mix adjustment models. As stated in the Introduction of this thesis, PROMs can measure a multitude of health aspects. Given this complexity of PROMs, it may be necessary that these concepts are quantified at baseline as case-mix factors especially because traditional case-mix factors may not entirely influence them. This issue is also related to the "ceiling effect" of PROMs: as baseline PROM scores improve, the size of any possible gain in scores diminishes in the follow-up measurements. For example, if one healthcare provider tends to treat patients with better baseline PROM scores (e.g. better quality of life), it may have a lower average gain than a healthcare provider that treats patients with lower baseline PROM scores. In relation to this issue, future research should also explore if there are differences in predictors in case-mix adjustment models between baseline and follow-up PROM scores outcomes.

Analysis of Prospectively Collected PROMs

If non-inferiority in terms of clinical outcomes has been established between treatment strategies, such as similar survival rates for early-stage breast cancer patients after either BCT or conventional mastectomy[54-56], it is very unlikely that future trials will be conducted for these patients with PROM scores as primary outcomes. Although ethical concerns would be minimal, the considerable price tag, limited generalizability, and long duration of an RCT would be valid reasons for not conducting one. Furthermore, observational PROMs data have the important advantage of reflecting the patient experience of actual health service delivery as opposed to a RCTs, which are often not representative of normal clinical practice. Therefore, mostly observational PROMs data will need to be relied on for drawing conclusions on the impact of interventions (treatments) on various domains of quality of life. Due to the increasing availability of PROMs data, there is a rising interest in quasi-experimental designs (e.g. regression discontinuity (RD) or instrumental variables (IV) analysis) to estimate treatment effectiveness. In the RD design, treatment allocation is not assigned randomly but rather based on a baseline assignment variable (e.g. age below or above 75 years) often related to the outcome of interest.[57] On the other hand, an IV-analysis is based on a variable that is highly associated with the exposure (or treatment) variable of interest but

not directly with the outcome of interest.[58] To date, there has been no literature on these quasi-experimental designs for PROMs data yet.

Several considerations in the analysis of PROMs need to be further explored to correct for the variability and statistical uncertainty inherent to observational data. Firstly, a PROM score at one collection time point represents a moment in time, a single snapshot of a patient's health status. Whether or not differences in PROM scores (delta's) are superior as quality measures to single PROM scores at follow-up time points remains an unanswered research question. By focusing only on post-treatment PROM scores as the outcome of interest, the opportunity to compare the before and after scores may be missed. Not only within one treatment modality, but also across two or more treatment strategies. Secondly, to enable the use of PROMs for individualized patient care, comparative effectiveness research and benchmarking, future endeavors should explore techniques to address the important issue of missingness within PROMs data. After all, missing data can lead to biased and misleading inferences on performance assessments. Complete case analysis assumes that missingness is independent of both observed and observed parameters of the data set (MCAR, "missing completely at random"), which previous studies have found an implausible assumption in the matter of PROMs.[57,58] Therefore, future research should focus more on (multiple) imputation techniques to handle missing PROMs data, assuming the differences between patients with complete or missing responses are entirely dependent on either measured values such as patient characteristics (MAR, "missing at random") or unmeasured values (MNAR, "missing not at random"). These techniques should not only take into the consideration observed factors related to the patient (e.g. stroke severity, breast cancer chemotherapy regimen) but also to observed factors related to the care provider (e.g. data collection method and frequency). Lastly, the correlations between different PROMs, both generic and disease-specific, has been insufficiently studied for most diseases. The strength of these correlations may alter over time, regardless of the disease, as the discriminatory potential of disease-specific and generic PROMs may be substantially different early after diagnosis, disease onset or treatment than over time. As some domains across different PROMs can overlap in terms of content and construct, future research on underlying relations is warranted if the desire is to reduce the registration burden for patients.

Minimal Clinically Important Differences (MCIDs)

Even though PROM score differences within (Chapter 6) or between breast cancer treatment strategies may exhibit statistical significance, this may not necessarily reflect a meaningful difference or change. These small score differences can be

influenced by not only small sample sizes, but by the unreliability of individual measurements.[59] One of the main reasons clinicians may be (understandably) skeptical about the usefulness of PROMs, is the scarcity of studies that have focused on improving the interpretation of PROM scores that clinicians would regard as meaningful. To increase the clinicians' familiarity and ability with interpreting PROM score differences, more research is required to establish reference scores, normative scores **(Chapter 7)** and MCIDs for the PROMs of interest.

The MCID can be defined as "the smallest difference in score in the domain of interest which *patients* perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's *management*".[39,60] This definition involves two constructs:
- a minimal amount of patient-reported score difference
- a difference significant enough to alter patient management

There are several methods to calculate MCIDs, but unfortunately these MCIDs can vary substantially depending on the method used. As a consequence, this has resulted in multiple methodological and interpretation problems.[61] Currently, there is no standard for calculating MCIDs but anchor-based methods have been described as superior to other methods.[62,63] This particular research field should be a high priority in order to encourage further uptake op PROMs in clinical care.

PERSPECTIVES ON FUTURE PRACTICAL APPLICATIONS OF PROMS IN CLINICAL CARE

Although patient-reported outcome measures were initially restricted to clinical trials, there is a growing use of PROMs in day-to-day clinical care for improving individual patient care.[64] PROMs can play multiple roles in the clinical care of patients. These include identifying patients' areas of concern that were previously unrecognized by care providers[41,65], monitoring patients' health status longitudinally[66], providing important information about treatment response prior or during clinic visits[67], improving the dialogue between patients and care providers in the consultation room[68], informing clinical decision-making[69] and promoting self-management in patients with chronic conditions[70]. If implemented well, PROMs can also potentially drive forward performance comparison (benchmarking individual care providers, teams, healthcare organizations) and quality-of-care improvement. This section contains important considerations that need to be

taken into account in current and future endeavors to implement and use PROMs in clinical care.

It is often the goal of healthcare professionals to use PROMs for multiple purposes and on several levels (micro-, meso- and macro-level). For instance, if PROMs are started in the consultation room in the context of individualized patient care (micro-level), professionals frequently already want to use these outcomes on an aggregated level for institutional benchmarking and quality improvement (meso-level) or even for population health monitoring and healthcare performance measurement (macro-level). This is an enormously challenging task as each aim sets different requirements for the questionnaire, time points for PROMs data administration/ collection, and statistical analysis. After all, for benchmarking it is essential that time points for PROMs data collection are standardized while for individualized care it is not. It is important for future health institutions to realize that the primary purpose of applying PROMs in clinical care should be the foundation for the entire implementation process. Other uses for PROMs data can be taken into account when setting up, but can also be scaled up towards at a later stage. For example, the primary goal of the routine collection of PROMs within the breast cancer care pathway in the Erasmus MC was to provide more individualized care to patients by enhancing patient-provider communication, prioritizing topics during the consultation, and improving shared decision-making. As the PROMs data increased over the past years following the start of the VBHC-initiative, clinicians also became interested in how this data may be used to gain insights in the trend of (health-related) quality of life scores within treatment strategies (Chapter 6). Future goals are to use the PROMs data to generate reference PROM scores at each collection time point to gain and give more perspective to the individual patient on their own score.

Prior to selecting PROMs as end points or outcomes, it is essential that the PROM's reliability, validity, and responsiveness have been tested for the patient population of interest. For example, as the sensitivity to change of generic PROMs may differ across acute (e.g. ischemic stroke, myocardial infarction) and chronic diseases (e.g. breast cancer, psoriasis), the instrument's responsiveness needs to be tested for various time points before selecting the time point at which PROMs are to be collected for benchmarking. Interestingly enough, prior research has shown that for the diseases addressed in this dissertation, namely ischemic stroke and breast cancer, the EQ-5D demonstrated an appropriate responsiveness within a year after disease onset.[71,72] Research into this important psychometric property should also inform decisions on the frequency of PROM administration when designing the

care pathway in which PROMs are prospectively collected. It is also important to realize that the longer the time period between the treatment and the outcome assessment, the higher the influence of confounding factors (e.g. age, gender, nationality, socio-economic status, life events) on the (patient-reported) outcomes rather than factors of the actual clinical care being provided.

Facilitators and barriers to the implementation of PROMs in clinical care can be categorized at individual, managerial and organizational levels.[44] The studies in Chapters 5 and 8 of this dissertation have both noted essential component in the PROM implementation process, namely the level of knowledge of care providers about the interpretability of PROM scores including scaling, psychometric properties and PROM-specific MCIDs. This lack of knowledge has a detrimental effect on the confidence and willingness of healthcare professionals to administer, review and discuss PROM results with their patients. Since many care providers have only recently begun to incorporate PROMs in their clinical care to, for instance, monitor treatment response, they may not share the same level of familiarity with the meaning of PROM (domain) scores or they may lack the ability to interpret variations in scores in contrast to traditional clinical markers (e.g. laboratory measures). Prior research has also shown that PROMs are valued by healthcare professionals if they prove to be useful for their clinical decision-making process.[41] Therefore, if care providers are engaged in the process of PROM collection and review, it will most likely positively influence the participation and adherence of patients themselves. By educating both patients and clinicians about the potential benefits of PROMs, it may potentially diminish the initial skepticism about their usefulness. As a result, the patient-provider communication in the consultation room may be facilitated as clinicians will be able to steer the conversation towards (previously unrecognized) health problems based on individual patients' responses and, if needed, be able to modify treatment goals based on patients' preferences.

Other implementation factors that surfaced in the aforementioned chapters were the IT-infrastructure and the clinical workflow. These factors are strongly linked as adequate IT-infrastructure can facilitate survey collection and PROMs data storage, leading to a reduced patient and clinician burden in completing or reviewing PROMs. Previous research has demonstrated that clinicians do not find PROMs disruptive to their workflow and actually find them useful.[73,74] This cooperation can be enhanced if more source code for PROMs software becomes modifiable and reusable for less experienced healthcare institutions. Although nine hospitals in the southwest region of the Netherlands had jointly decided to measure a standardized breast cancer outcomes set for possible benchmarking, there still seems

to be insufficient sharing of specialized technological and methodological expertise to measure those (patient-reported) outcomes. As a consequence, hospitals needlessly attempt to reinvent the technological wheel leading to fragmentation of efforts. Considering the substantial costs and efforts to develop electronic PROM data collection and assessment software, it is essential that healthcare institutions cooperate by sharing knowledge and experience on IT-infrastructure components (from integration in the EHR to a user-friendly dashboard for PROM score evaluation). Lastly, the collection, storage and management of PROMs data also requires staff. While nurse practitioners and administrative personnel have traditionally undertaken these responsibilities, the creation of a PROM coordinator position may be more beneficial in being the point of contact between clinicians, nurse practitioners, patients and IT-staff.

Box 9.2 | Overview of Key Recommendations

Part I: Methodological Aspects of PROMs

Patient-Reported Morbidity

- Patient-reported morbidity instruments should be validated for the population and outcome of interest (clinical, patient-reported or healthcare utilization)
- Developers of future morbidity instruments should, with regard to the item-level concordance with other data sources, preferably include diseases clearly defined within the vernacular, and diseases that require ongoing therapy (e.g. chronic conditions)

Case-Mix Adjustment for PROMS

- Case-mix adjustment is most useful when patient characteristics vary substantially between treatments or healthcare providers, and where these patient characteristics are strongly related to PROMs
- Before the selection of case-mix variables for PROMs, the goal of the case-mix adjustment model (internal vs. external benchmarking) needs to be decided
- Non-traditional case-mix factors (e.g. socio-economic status, ethnicity, psychological and social status) should be considered for case-mix adjustment models for PROMs as opposed to clinical outcomes
- Baseline PROM scores should be considered for case-mix adjustment models unless the outcome of interest is the difference in PROM scores (delta's)

Part II: Practical Applications in Clinical Care

Use of (Prospectively) Collected PROMs

- Establishing a research unit, comprising of epidemiologists, clinicians and psychologists, is essential for monitoring data collection and supporting data analysis
- PROMs need to be linked to clinical data (including quality indicators) on an aggregated level to determine the "full" effectiveness of treatment modalities
- Standardization of PRO instruments and data collection time points is required for external benchmarking (both on national and international level)

Implementation Practical Application of PROMs

- Before implementing PROMs, one primary purpose of PROMs needs to be chosen. After the routine use has been established, the primary purpose can be scaled up to additional purposes.
- Trained personnel (including physicians, nurses, administrative and IT-staff) is required to coordinate and support PROMs data collection, and to review and ultimately act upon PROM scores
- High response rates need to be prioritized through reducing registration burden for patients by aligning PROMs over follow-up and allowing sufficient time for survey administration
- Constant evaluation and optimization of PRO instruments, data collection methods and frequencies is required to improve the usability of PROMs data

FINAL REMARKS

This dissertation aimed to expand the evidence base on both methodological aspects and practical applications of patient-reported outcome measures, measures that enable the placement of patients at the heart of their own care.

As case-mix adjustment reduces bias and improves the validity of performance measurement, it not only facilitates quality of care improvement within hospitals but also enables meaningful and accurate comparisons between healthcare providers. Specific modifications to the case-mix adjustment methodology for patient-reported outcome measures need to be applied, depending on the goal of the adjustment (internal vs. external benchmarking). These modifications should be based on statistical models which predict PROMs, taking patient characteristics and other factors into account which are beyond the control of providers. By doing so, case-mix will better assist organizations in evaluating healthcare services.

Current outcomes are still limited and have a tendency to focus on measures of failure (e.g. treatment complications) rather than measures of success (e.g. improvement in health-related quality of life). To ensure an increase in the adoption of PROMs in clinical care, more studies on the clinical implications (e.g. effect of routine PRO measurements on clinical outcomes, patient-provider communication, and health service utilization) should be performed. The importance of this should be even more emphasized given the rise of chronic conditions that need to be managed with a holistic approach.

REFERENCES

1. Health USDo, Human Services FDACfDE, Research, et al. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes.* 2006;4:79.
2. Black N. Patient reported outcome measures could help transform healthcare. *Bmj.* 2013;346:f167.
3. Sarfati D. Review of methods used to measure comorbidity in cancer populations: No gold standard exists. *Journal of Clinical Epidemiology.* 2012;65(9):924-933.
4. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. *Arthritis Rheum.* 2003;49(2):156-163.
5. Preen DB, Holman CDAJ, Lawrence DM, Baynham NJ, Semmens JB. Hospital chart review provided more accurate comorbidity information than data from a general practitioner survey or an administrative database. *Journal of Clinical Epidemiology.* 2004;57(12):1295-1304.
6. Januel JM, Luthi JC, Quan H, et al. Improved accuracy of co-morbidity coding over time after the introduction of ICD-10 administrative data. *BMC health services research.* 2011;11:194.
7. Klabunde CN, Warren JL, Legler JM. Assessing comorbidity using claims data: an overview. *Medical care.* 2002;40(8 Suppl):IV-26-35.
8. Sarfati D, Gurney J, Stanley J, Lim BT, McSherry C. Development of a pharmacy-based comorbidity index for patients with cancer. *Medical care.* 2014;52(7):586-593.
9. de Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity. a critical review of available methods. *J Clin Epidemiol.* 2003;56(3):221-229.
10. Valderas JM, Starfield B, Sibbald B, Salisbury C, Roland M. Defining comorbidity: implications for understanding health and health services. *Ann Fam Med.* 2009;7(4):357-363.
11. Schellevis FG, van der Velden J, van de Lisdonk E, van Eijk JTM, van Weel C. Comorbidity of chronic diseases in general practice. *Journal of Clinical Epidemiology.* 1993;46(5):469-473.
12. Ording AG, Sørensen HT. Concepts of comorbidities, multiple morbidities, complications, and their clinical epidemiologic analogs. *Clin Epidemiol.* 2013;5:199-203.
13. Katz JN, Chang LC, Sangha O, Fossel AH, Bates DW. Can comorbidity be measured by questionnaire rather than medical record review? *Med Care.* 1996;34(1):73-84.
14. Bayliss EA, Ellis JL, Steiner JF. Subjective assessments of comorbidity correlate with quality of life health outcomes: initial validation of a comorbidity assessment instrument. *Health Qual Life Outcomes.* 2005;3:51.
15. Pettengill J, Vertrees J. Reliability and validity in hospital case-mix measurement. *Health Care Financing Review.* 1982;4(2):101.
16. o Malley AJ, Zaslavsky A, Elliott M, Zaborski L, Cleary P. Case-mix adjustment of the CAHPS Hospital Survey. *Health services research.* 2005;40 6 Pt 2:2162-2181.
17. Iezzoni LI. 3. 1 Risk adjustment for performance measurement. *Performance measurement for health system improvement: experiences, challenges and prospects.* 2009;251.
18. Fischer C, Lingsma H, Hardwick R, Cromwell DA, Steyerberg E, Groene O. Risk adjustment models for short-term outcomes after surgical resection for oesophagogastric cancer. *BJS (British Journal of Surgery).* 2016;103(1):105-116.
19. Talsma AK, Reedijk AMJ, Damhuis RAM, Westenend PJ, Vles WJ. Re-resection rates after breast-conserving surgery as a performance indicator: introduction of a case-mix model

to allow comparison between Dutch hospitals. *European Journal of Surgical Oncology (EJSO).* 2011;37(4):357-363.
20. Teale EA, Forster A, Munyombwe T, Young JB. A systematic review of case-mix adjustment models for stroke. *Clinical rehabilitation.* 2012;26(9):771-786.
21. *Patient Reported Outcome Measures (PROMs) in England: The Case-Mix Adjustment Methodology.* London: Department of Health;2012.
22. Burgess R, Bishop A, Lewis M, Hill J. Models used for case-mix adjustment of patient reported outcome measures (PROMs) in musculoskeletal healthcare: A systematic review of the literature. *Physiotherapy.* 2018;105.
23. Appleby J, Devlin N, Parkin D. *Using patient reported outcomes to improve health care.* John Wiley & Sons; 2015.
24. Berkhemer OA, Fransen PS, Beumer D, et al. A randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med.* 2015;372(1):11-20.
25. Gibbons E, Black N, Fallowfield L, Newhouse R, Fitzpatrick R. Patient-reported outcome measures and the evaluation of services. *Challenges, solutions and future directions in the evaluation of service innovations in health care and public health*: NIHR Journals Library; 2016.
26. Santana MJ, Haverman L, Absolom K, et al. Training clinicians in how to use patient-reported outcome measures in routine clinical practice. *Quality of Life Research.* 2015;24(7):1707-1718.
27. Weldring T, Smith SMS. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health Serv Insights.* 2013;6:61-68.
28. Jensen RE, Rothrock NE, DeWitt EM, et al. The role of technical advances in the adoption and integration of patient-reported outcomes in clinical care. *Medical care.* 2015;53(2):153-159.
29. Snyder CF, Aaronson NK, Choucair AK, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. *Quality of Life Research.* 2012;21(8):1305-1314.
30. van Egdom LSE, Lagendijk M, van der Kemp MH, et al. Implementation of Value Based Breast Cancer Care. *European Journal of Surgical Oncology.* 2019;45(7):1163-1170.
31. Ong WL, Schouwenburg MG, Van Bommel ACM, et al. A standard set of value-based patient-centered outcomes for breast cancer: the international consortium for health outcomes measurement (ICHOM) initiative. *JAMA oncology.* 2017;3(5):677-685.
32. Dronkers EAC, Baatenburg de Jong RJ, van der Poel EF, Sewnaik A, Offerman MPJ. Keys to successful implementation of routine symptom monitoring in head and neck oncology with "Healthcare Monitor" and patients' perspectives of quality of care. *Head & neck.* 2020.
33. Melis-Schrijver LM, den Hollander-Ardon MS, Janssen WG, van der Veen R. Gemstracker Expertise team Rotterdam arm and hand: Web-based monitoring of physical, social and emotional functioning in patients with upper limb amputations. *Prosthet Orthot Int.* 2018;42(1):50-55.
34. Sjoding MW, Luo K, Miller MA, Iwashyna TJ. When do confounding by indication and inadequate risk adjustment bias critical care studies? A simulation study. *Crit Care.* 2015;19(1):195.
35. Sørensen HT, Lash TL, Rothman KJ. Beyond randomized controlled trials: a critical comparison of trials with nonrandomized studies. *Hepatology.* 2006;44(5):1075-1082.

36. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-World Evidence — What Is It and What Can It Tell Us? *New England Journal of Medicine.* 2016;375(23):2293-2297.
37. Mundy LR, Homa K, Klassen AF, Pusic AL, Kerrigan CL. Breast Cancer and Reconstruction: Normative Data for Interpreting the BREAST-Q. *Plast Reconstr Surg.* 2017;139(5):1046e-1055e.
38. Klifto KM, Aravind P, Major M, et al. Establishing Institution-Specific Normative Data for the BREAST-Q Reconstruction Module: A Prospective Study. *Aesthet Surg J.* 2020;40(6):NP348-NP355.
39. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Controlled clinical trials.* 1989;10(4):407-415.
40. Lagendijk M, van Egdom LSE, Richel C, et al. Patient reported outcome measures in breast cancer patients. *Eur J Surg Oncol.* 2018;44(7):963-968.
41. Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. *BMJ quality & safety.* 2014;23(6):508-518.
42. Foster A, Croot L, Brazier J, Harris J, O'Cathain A. The facilitators and barriers to implementing patient reported outcome measures in organisations delivering health related services: a systematic review of reviews. *J Patient Rep Outcomes.* 2018;2:46-46.
43. Philpot LM, Barnes SA, Brown RM, et al. Barriers and Benefits to the Use of Patient-Reported Outcome Measures in Routine Clinical Care: A Qualitative Study. *Am J Med Qual.* 2018;33(4):359-364.
44. Duncan EAS, Murray J. The barriers and facilitators to routine outcome measurement by allied health professionals in practice: a systematic review. *BMC health services research.* 2012;12(1):1-9.
45. Antunes B, Harding R, Higginson IJ, Euroimpact. Implementing patient-reported outcome measures in palliative care clinical practice: a systematic review of facilitators and barriers. *Palliative medicine.* 2014;28(2):158-175.
46. Sørensen K, Pelikan JM, Röthlin F, et al. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). *Eur J Public Health.* 2015;25(6):1053-1058.
47. Habbous S, Chu KP, Harland LT, et al. Validation of a one-page patient-reported Charlson comorbidity index questionnaire for upper aerodigestive tract cancer patients. *Oral Oncol.* 2013;49(5):407-412.
48. Simpson CF, Boyd CM, Carlson MC, Griswold ME, Guralnik JM, Fried LP. Agreement between self-report of disease diagnoses and medical record validation in disabled older women: factors that modify agreement. *J Am Geriatr Soc.* 2004;52(1):123-127.
49. Iecovich E, Biderman A. Concordance between self-reported and physician-reported chronic co-morbidity among disabled older adults. *Can J Aging.* 2013;32(3):287-297.
50. Van Der Wees PJ, Nijhuis-Van Der Sanden MW, Ayanian JZ, Black N, Westert GP, Schneider EC. Integrating the use of patient-reported outcomes for both clinical practice and performance measurement: views of experts from 3 countries. *Milbank Q.* 2014;92(4):754-775.
51. Paddison C, Elliott M, Parker R, et al. Should measures of patient experience in primary care be adjusted for case mix? Evidence from the English General Practice Patient Survey. *BMJ quality & safety.* 2012;21(8):634-640.
52. Berwick D, Black N, Cullen D, et al. Recommendations to OECD ministers of health from the high level reflection group on the future of health statistics: Strengthening the inter-

national comparison of health system performance through patient-reported indicators. *Statistics.* 2017.
53. Vranceanu AM, Bachoura A, Weening A, Vrahas M, Smith RM, Ring D. Psychological factors predict disability and pain intensity after skeletal trauma. *J Bone Joint Surg Am.* 2014;96(3):e20.
54. van Maaren MC, de Munck L, de Bock GH, et al. 10 year survival after breast-conserving surgery plus radiotherapy compared with mastectomy in early breast cancer in the Netherlands: a population-based study. *The Lancet Oncology.* 2016;17(8):1158-1170.
55. Lagendijk M, van Maaren MC, Saadatmand S, et al. Breast conserving therapy and mastectomy revisited: Breast cancer-specific survival and the influence of prognostic factors in 129,692 patients. *Int J Cancer.* 2018;142(1):165-175.
56. Litiere S, Werutsky G, Fentiman IS, et al. Breast conserving therapy versus mastectomy for stage I-II breast cancer: 20 year follow-up of the EORTC 10801 phase 3 randomised trial. *Lancet Oncol.* 2012;13(4):412-419.
57. van Leeuwen N. *Design and Analysis of Randomized and Non-randomized Studies: Improving Validity and Reliability* [Ph.D. thesis], Erasmus University Rotterdam; 2019.
58. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *Journal of clinical epidemiology.* 2009;62(12):1226-1232.
59. Hutchings A, Neuburger J, van der Meulen J, Black N. Estimating recruitment rates for routine use of patient reported outcome measures and the impact on provider comparisons. *BMC health services research.* 2014;14:66.
60. Gutacker N, Street A, Gomes M, Bojke C. Should English healthcare providers be penalised for failing to collect patient-reported outcome measures? A retrospective analysis. *J R Soc Med.* 2015;108(8):304-316.
61. McDowell I. *Measuring health: a guide to rating scales and questionnaires.* Oxford University Press, USA; 2006.
62. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol.* 2002;14(2):109-114.
63. Hajiro T, Nishimura K. Minimal clinically significant difference in health status: the thorny path of health status measures? *Eur Respir J.* 2002;19(3):390-391.
64. Johnston BC, Ebrahim S, Carrasco-Labra A, et al. Minimally important difference estimates and methods: a protocol. *BMJ open.* 2015;5(10).
65. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of clinical epidemiology.* 2008;61(2):102-109.
66. Field J, Holmes MM, Newell D. PROMs data: can it be used to make decisions for individual patients? A narrative review. *Patient Relat Outcome Meas.* 2019;10:233-241.
67. Valderas JM, Kotzeva A, Espallargues M, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Quality of Life Research.* 2008;17(2):179-193.
68. Espallargues M, Valderas JM, Alonso J. Provision of feedback on perceived health status to health care professionals: a systematic review of its impact. *Medical care.* 2000;38(2):175-186.

69. Holmes MM, Bishop FL, Newell D, Field J, Lewith G. Chiropractors' views on the use of patient-reported outcome measures in clinical practice: a qualitative study. *Chiropr Man Therap.* 2018;26:50.
70. Boyce MB, Browne JP. Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients? A systematic review. *Qual Life Res.* 2013;22(9):2265-2278.
71. Tyson SF, Greenhalgh J, Long AF, Flynn R. The influence of objective measurement tools on communication and clinical decision making in neurological rehabilitation. *J Eval Clin Pract.* 2012;18(2):216-224.
72. Santana MJ, Feeny D. Framework to assess the effects of using patient-reported outcome measures in chronic care management. *Qual Life Res.* 2014;23(5):1505-1513.
73. Golicki D, Niewada M, Karlińska A, et al. Comparing responsiveness of the EQ-5D-5L, EQ-5D-3L and EQ VAS in stroke patients. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2015;24(6):1555-1563.
74. Kimman ML, Dirksen CD, Lambin P, Boersma LJ. Responsiveness of the EQ-5D in breast cancer patients in their first year after treatment. *Health Qual Life Outcomes.* 2009;7:11.
75. Kelleher SA, Somers TJ, Locklear T, Crosswell AD, Abernethy AP. Using patient reported outcomes in oncology clinical practice. *Scandinavian journal of pain.* 2016;13:6-11.
76. Jensen RE, Snyder CF, Abernethy AP, et al. Review of electronic patient-reported outcomes systems used in cancer clinical care. *Journal of Oncology Practice.* 2014;10(4):e215-e222.

Chapter 10 Summary

GENERAL INTRODUCTION

One of the components of Value-based Healthcare (VBHC) is the measurement of outcomes that matter to patients. Patient-reported outcome measures (PROMs) are the instruments to capture outcomes directly reported by patients on his or her own health status (including quality of life) and/ or treatment, thus without external interpretation of patients' responses by healthcare professionals. These measures express the combined influence of diseases or health conditions and their treatments. PROMs have been implemented in clinical care with wide-ranging intentions: from screening of health problems or psychosocial issues within primary care, monitoring patients' response to treatment, facilitating patient-provider communication, enhancing shared decision-making, to benchmarking healthcare institutions for quality of care improvement and transparency.

The aim of this dissertation was two-fold: to study certain methodological aspects of the analysis (e.g. important case-mix factors to consider and how to adjust for them) of PROMs (Part I; Chapters 2–4), and to explore the implementation and practical application of PROMs in clinical care (Part II; Chapters 5–8).

Specific research questions in this thesis were:

1. Which general patient-reported morbidity instruments (in order to use for case-mix adjustment) have been published? (*Chapter 2*)
2. a) Which predictors are important for case-mix adjustment models for a generic patient-reported outcome measure (EQ-5D) within acute ischemic stroke care? (*Chapter 3*)
   b) How do case-mix adjustment models for the EQ-5D differ from models for clinical outcomes (mortality, modified Rankin Scale) within acute ischemic stroke care? (*Chapter 3*)
3. a) What are the differences in case-mix adjustment models between individual domains of a generic patient-reported outcome measure (EQ-5D) within acute ischemic stroke? (*Chapter 4*)
   b) What is the influence of case-mix adjustment on the reliability of individual EQ-5D domains as quality indicators for between-hospital comparisons in the quality of ischemic stroke care? (*Chapter 4*)
4. a) How have patient-reported outcome measures been administered within clinical breast cancer care? (*Chapter 5*)
   b) What is the impact of (routine) PROM collection on patients, care providers, and care processes? (*Chapter 5*)

   c) Which facilitators and barriers of PROM collection have been reported? (*Chapter 5*)
5. What is the course of PROM scores over time within (surgical) treatment strategies for breast cancer? (*Chapter 6*)
6. What are the normative BREAST-Q scores in a sample of Dutch women without a prior history of breast cancer? (*Chapter 7*)
7. What are the facilitators and barriers for implementing PROMs in clinical care, as reported by healthcare professionals of an academic center? (*Chapter 8*)

PART I – Methodological Aspects of PROMs

The first part of this thesis addresses the methodological aspects of measuring analyzing patient-reported outcome measures (PROMs), including an overview of instruments for capturing patient-reported morbidity (as a case-mix factor), and an exploration of the development of case-mix adjustment models specifically for PROMs .

In **Chapter 2**, a systematic review on general patient-reported morbidity measures is presented. In this review, a total of 34 studies were included which described the development and/ or psychometric performance of instruments. The Self-Reported Charlson Comorbidity Index (1996), the Self-Administered Comorbidity Questionnaire (2003), and the Disease Burden Morbidity Assessment (2005) were the most frequently assessed patient-reported morbidity instruments in the literature. Most often included diseases/ conditions across the instruments were diabetes, hypertension and myocardial infarction, with diabetes having the highest item-level reliability (based on Kappa (κ) coefficients) across most included studies. A considerable range in item-level reliability of instruments versus the gold standard medical record review (κ range 0.66 – 0.86) was observed within studies, suggesting that the concordance is dependent on the specific disease/ condition in question. Only a few studies demonstrated the level of concordance of morbidity instruments with other data sources: the agreement with medical records was moderate, while the agreement with administrative data was notably poor. Furthermore, the associations of a morbidity instrument's score with various outcomes of healthcare utilization (e.g. hospitalizations, hospital stay, emergency room visits) were analyzed in only 7 studies, which showed mostly weak relations.

Chapter 3 described a study which aimed to develop and compare case-mix adjustment models for two clinical (mortality and modified Rankin Scale (mRS)) outcomes and a patient-reported outcome. A subset of data from the Dutch National Stroke Registry was used, which included ischemic stroke patients treated

in four stroke centers who had complete mRS and EQ-5D data three months post-discharge (N = 1022), after multiple imputation was performed 10 times to replace missing baseline patient characteristics (among which case-mix variables). The models for mortality, mRS score and EQ-5D index score were developed with respectively logistic, ordinal and linear regression models with stepwise backward selection. Exclusive predictors for the EQ-5D index score were socio-economic status, sex, and nationality. In comparison to the models for mortality (R^2 = 0.44) and mRS (R^2 = 0.42), the linear regression model for the EQ-5D had the lowest goodness-of-fit (R^2 = 0.37). The largest increase in R^2 of the models for mortality and mRS was observed after adding NIHSS, while for the EQ-5D index score it was after the addition of age as a predictor. After mRS and EQ-5D index scores were both dichotomized as outcomes, AUC's were compared: AUC = 0.87 (mortality) vs. AUC = 0.83 (mRS ≥ 3) vs. AUC = 0.78 (EQ-5D index score ≥ 0.65). In contrast to the models for mortality and mRS, more predictors were needed in the model for the EQ-5D index score for the predictive ability to reach a plateau.

Building on these results, the aim of the study in **Chapter 4** was to develop case-mix adjustment models for individual EQ-5D domain scores after acute ischemic stroke and to compare hospital effect estimates for these scores before and after case-mix adjustment. A secondary analysis was performed on a data subset (N = 336) from the MR CLEAN study, after multiple imputation (10 iterations) was performed to replace missing values of candidate predictors and the exclusion of four stroke centers that included less than 5 patients, patients who died before the three-month follow-up, and patients with missing EQ-5D data. In this substudy, the severity responses to each individual EQ-5D domain were used as outcomes for five individual ordinal regression models. Common predictors in the regression models across all individual EQ-5D domains were hypertension and peripheral arterial occlusive disease. R^2 values ranged from 0.31 ("self-care"), 0.27 ("mobility"), 0.19 ("usual activities"), 0.13 ("pain/discomfort"), to 0.08 ("anxiety/depression"). After case-mix adjustment, estimated performance of individual stroke centers changed only slightly for "physical" domains as opposed to "psychological" domains.

PART II – PRACTICAL APPLICATIONS OF PROMS IN CLINICAL CARE

The second part of this dissertation not only explored the implementation of PROMs in clinical care including the facilitators and barriers perceived or experi-

enced by healthcare professionals, but the prospectively collected PROMs of both healthy women and breast cancer patients as well.

The systematic review in **Chapter 5** provided an overview of 34 articles about PROM interventions (including collection/ administration methods and frequencies) in breast cancer care, its facilitators and barriers, and its impact on patients, clinicians and care processes. The most frequently used instruments across articles were the CTCAE, the EORTC-QLQ-C30 and the EQ-5D. Across most included studies, the administration of PROMs was during treatment with PROMs being mostly collected electronically through web-based assessments or tablets. A promising effect of PROM collection on patients (e.g. symptom distress, quality of life, and satisfaction), providers (e.g. willingness to comply, symptom management), and care process or system outcomes (e.g. patient-provider communication, hospital visits). Most reported facilitators and barriers were mostly of a technical nature (e.g. IT-infrastructure).

In **Chapter 6,** an analysis was conducted on postoperative course of PROM scores of breast cancer patients who underwent either breast-conserving therapy (BCT) (N = 143) and mastectomy without reconstruction (N = 70) in the Erasmus University Medical Center. Within both surgical strategies, PROMs (BREAST-Q, EORTC-QLQ-C30 and EORTC-QLQ-BR23) were collected and compared from baseline (prior to surgery) till 12 months post-primary surgery. Multiple imputation was used to handle the missing PROM scores at either T0, T6 and/ or T12. At 12 months follow-up, there was a significant improvement in “Body Image” (EORTC-QLQ-BR23) for patients following BCT, while “Physical Wellbeing” (BREAST-Q), “Physical Functioning” (EORTC-QLQ-C30), “Sexual Wellbeing” (BREAST-Q) all significantly declined. A statistically significant reduction in both median “Satisfaction with Breasts” (BREAST-Q) and “Psychosocial Wellbeing” (BREAST-Q) scores was observed for patients who underwent mastectomy without reconstruction. In addition, there was a moderately strong correlation ($r = –0.75$, $p < 0.001$) observed between “Physical Wellbeing” and “Breast Symptoms” (EORTC-QLQ-BR23) at T12 for patients following BCT, while “Satisfaction with Breasts” and “Body Image” was moderately correlated ($r = 0.68$, $p < 0.001$) at T12 for patients who underwent mastectomy. A comparison between respondents and non-respondents at baseline (T0) across all three included PROMs did not reveal significantly different patient characteristics.

Chapter 7 described a cross-sectional study which determined Dutch normative BREAST-Q values, in an effort to provide more context for both patients and clini-

cians when interpreting breast cancer patients' scores. Women were invited to self-complete a web-based survey, including demographic questions and a Dutch translated version of the BREAST-Q. In total, questionnaires from 9059 Dutch women, without a prior history of breast cancer, were analyzed. Mean normative scores for BREAST-Q domains were 64.24 ± SD 18.60 ("Satisfaction with Breasts"), 71.95 ± SD 15.93 ("Psychosocial Wellbeing"), 89.54 ± SD 12.48 ("Physical Wellbeing") and 60.38 ± SD 15.37 ("Sexual Wellbeing"). Notably, Dutch normative BREAST-Q scores seem to follow similar patterns across the domains in comparison to two previously published BREAST-Q normative data from USA. Prior non-cancer-related breast surgery was associated with higher "Satisfaction with Breasts" and "Sexual Wellbeing" scores, but had a negative association with "Physical Wellbeing".

The cross-sectional study in **Chapter 8** on healthcare professionals' experience and perspectives on the implementation of PROMs was sparked by a further interest in its facilitators and barriers after the systematic review, described in Chapter 5, was conducted**.** A customized questionnaire was developed, containing both statements (with a 5-point Likert scale) and 8 open-ended questions on perceived or experienced facilitators and barriers of implementing PROMs in clinical care. 61 respondents (response rate 33%) in various stages of the implementation process were included, who were mostly medical specialists (51%) and had a median employment duration of 14 years. Most participants were involved in patient care for Turner syndrome, cleft lip and palate and gynecological cancers. Facilitators, that were most often mentioned by study participants, included the presence of a coordinator in the clinic and the multi-disciplinary collaboration. Frequently reported barriers were the challenging IT-infrastructure (including access to the web-based PROM platform, dashboard visualization, integration of PROMs in the electronic health record) and time constraints in combination with the labor-intensive nature of the incorporation of PROMs in care.

GENERAL DISCUSSION

The aim of this dissertation was to study aspects of both the analysis (including case-mix adjustment models) and practical application (including implementation) of PROMs in clinical care. The findings in this dissertation have implications for both research and clinical care.

In the first half of this thesis, case-mix adjustment models for a generic patient-reported outcome measure (EQ-5D) for ischemic stroke patients were developed

for either comparison to models for clinical outcomes or to illustrate important caveats of benchmarking PROMs. With regard to case-mix, the model performance for the generic PROM was considerably less compared to the models for clinical outcomes, with more predictors being required for the EQ-5D model in order for its predictive ability to reach a plateau. The explanation may be found in the lack of additional non-traditional baseline predictors (e.g. psychological, social and quality of life factors), which may contribute more to model performances for "subjective" outcomes like PROMs. In addition, it was demonstrated that model performances for individual (psychological) EQ-5D domains "anxiety/ depression" and "pain/ discomfort" were considerably lower than for the (physical) domains of "usual activities", "mobility" and "self-care". These findings also suggest that PROMs are outcomes that are more influenced by a wide variety of case-mix factors than traditional (clinical) outcomes. Therefore, if the desire is to use PROMs for benchmarking, then the collection of non-traditional case-mix factors will be necessary. An important research aim for the future is the identification of case-mix factors that are strongly associated with frequently used PROMs and to assess the differences in model performances of case-mix adjustment models between clinical outcomes and PROMs.

The second half of this dissertation has focused on the practical application of PROMs in clinical care including methods for administration, the analysis of longitudinally collected PROMs, the collection of normative PROM data, and the identification of facilitators and barriers of implementing PROMs in clinical care as perceived by healthcare professionals. PROMs have the potential to improve the quality of care by addressing issues of importance to the patient. The longitudinal collection of PROMs can, however, be very challenging due to issues revolving around IT-infrastructure, compliance over time, and clinicians' lack of time or knowledge. The uptake of PROMs in clinical care is more likely to succeed with a multi-disciplinary approach, which involves physicians, nurses, patients, management, researchers, data manager, IT- and administrative personnel.

To promote individualized care, patients need to be informed by care providers about the objectives of the PROMs in order to build trust and compliance. In relation to this, normative data can play an important role to contextualize the PROM scores of patients and improve patient-provider communication. Through continued research and engagement of healthcare professionals in sharing expertise on the application of PROMs in clinical care, the enhancement of care delivery through PROMs will continue to evolve.

Chapter 10 Samenvatting

INTRODUCTIE

Eén van de componenten van Value-Based Healthcare (VBHC) of Waardegedreven Zorg is het meten van uitkomsten en het sturen hierop. Patiënt-gerapporteerde uitkomstmaten (PROMs) zijn instrumenten die de zelfervaren gezondheidsstatus (inclusief symptomen, kwaliteit van leven, functioneren in dagelijkse activiteiten) van een patiënt kunnen meten en weergeven, zonder interpretatie door een zorgprofessional. Deze vragenlijsten kunnen ook informatie verschaffen over hoe de patiënt zijn of haar zorgproces ervaart. PROMs worden geïmplementeerd in de klinische praktijk met een grote variatie van doeleinden: van het screenen van gezondheidsklachten en psychosociale problemen binnen de primaire zorg, het monitoren van de respons van de behandeling op gezondheid en kwaliteit van leven, het faciliteren van de communicatie tussen patiënt en zorgverlener, het verbeteren van gedeelde besluitvorming, tot het benchmarken van zorginstellingen voor kwaliteitsverbetering en transparantie.

Dit proefschrift heeft twee doelen: het ontwikkelen en vergelijken van case-mix correctiemodellen voor zowel klinische uitkomsten en PROMs (Deel I; Hoofdstukken 2–4), en het vergroten van de wetenschappelijke kennis over de implementatie en praktische toepassing van PROMs in de klinische zorg (Deel II; Hoofdstukken 5–8).

De specifieke onderzoeksvragen in dit proefschrift zijn:

1. Welke algemene patiënt-gerapporteerde morbiditeitsinstrument (ter gebruik voor case-mix correctie) zijn gepubliceerd? (*Hoofdstuk 2*)
2. a) Welke voorspellers zijn belangrijk in een case-mix correctie model voor een generieke patiënt-gerapporteerde uitkomstmaat (EQ-5D) binnen de ischemische beroertezorg? (*Hoofdstuk 3*)
   b) In hoeverre verschilt een case-mix correctie model voor de EQ-5D ten opzichte van klinische uitkomsten (mortaliteit, modified Rankin Scale) binnen de ischemische beroertezorg? (*Hoofdstuk 3*)
3. a) Wat zijn de verschillen in case-mix correctiemodellen tussen individuele domeinen van een generieke patiënt-gerapporteerde uitkomstmaat (EQ-5D) binnen de ischemische beroertezorg? (*Hoofdstuk 4*)
   b) Wat is de invloed van case-mix correctie op de betrouwbaarheid van individuele EQ-5D domeinen als kwaliteitsindicatoren voor ziekenhuisvergelijkingen in de kwaliteit van de ischemische beroertezorg? (*Hoofdstuk 4*)
4. a) Hoe zijn patiënt-gerapporteerde uitkomstmaten afgenomen in de klinische borstkankerzorg? (*Hoofdstuk 5*)

   b) Wat is de impact van (routinematige) PROM collectie op patiënten, zorgverleners, en zorgprocessen? (*Hoofdstuk 5*)
   c) Welke facilitators en barrières van PROM collectie zijn gerapporteerd? (*Hoofdstuk 5*)
5. Wat is het beloop van PROM scores over de tijd binnen (chirurgische) behandelstrategieën van borstkanker? (*Hoofdstuk 6*)
6. What zijn de normatieve BREAST-Q scores van Nederlandse vrouwen zonder borstkanker in de voorgeschiedenis? (*Hoofdstuk 7*)
7. Wat zijn, volgens medische professionals van een academisch ziekenhuis, de facilitators en barrières van het implementeren van PROMs in de klinische praktijk? (*Hoofdstuk 8*)

DEEL I – METHODOLOGISCHE ASPECTEN VAN PROMS

Het eerste deel van dit proefschrift behandelt de methodologische aspecten van de analyse van patiënt-gerapporteerde uitkomstmaten (PROMs), inclusief een overzicht van instrumenten voor het vastleggen patiënt-gerapporteerde morbiditeiten (als een case-mix factor), en een verkenning in de ontwikkeling van case-mix correctiemodellen gericht op PROMs.

Hoofdstuk 2 beschrijft een systematische review van algemene patiënt-gerapporteerde morbiditeitsvragenlijsten. In totaal werden 34 studies geïncludeerd die de ontwikkeling en/ of psychometrische eigenschappen van instrumenten beschreven. De Self-Reported Charlson Comorbidity Index (1996), de Self-Administered Comorbidity Questionnaire (2003) en de Disease Burden Morbidity Assessment (2005) waren de meest beoordeelde patiënt-gerapporteerde morbiditeitsvragenlijsten in de literatuur. De meest voorkomende ziekten/ aandoeningen in de instrumenten waren diabetes, hypertensie en myocardinfarct, waarbij diabetes de hoogste betrouwbaarheid op itemniveau (gebaseerd op Kappa (κ)-coëfficiënten) had. Een aanzienlijk verschil in betrouwbaarheid op itemniveau tussen vragenlijsten versus medische dossiers werd waargenomen in studies, suggererend dat de mate van overeenstemming afhankelijk is van de specifieke ziekte/ aandoening in kwestie. Slechts enkele studies toonden de mate van overeenstemming aan van morbiditeitsinstrumenten met andere databronnen: de overeenstemming met medische dossiers was matig, terwijl die met administratieve data opmerkelijk slecht was. Bovendien werden de associaties tussen de scores van morbiditeitsvragenlijsten en verschillende maten van zorggebruik (bijv. ziekenhuisopnames

en -verblijf, spoedeisende hulp bezoeken) in slechts 7 studies geanalyseerd, die voornamelijk zwakke verbanden lieten zien.

Hoofdstuk 3 beschrijft een studie die gericht was op het ontwikkelen en vergelijken van case-mix aanpassingsmodellen voor twee klinische uitkomsten (modified Rankin Scale (mRS) en mortaliteit) en een patiënt-gerapporteerde uitkomstmaat (EQ-5D). Een subset van data uit het Nederlandse CVA register werd gebruikt, die patiënten met een ischemische beroerte omvatte die werden behandeld in vier centra voor beroerte en die drie maanden na ontslag complete mRS en EQ-5D data hadden (N = 1022), nadat 10 keer multiple imputation werd uitgevoerd ter vervanging van missing baseline patiëntkarakteristieken (waaronder case-mix variabelen). De modellen voor mortaliteit, de mRS score en EQ-5D-index score werden ontwikkeld met respectievelijk logistieke, ordinale en lineaire regressiemodellen met "stepwise backward selection". Exclusieve voorspellers voor de EQ-5D score waren sociaaleconomische status, geslacht en nationaliteit. In vergelijking met de modellen voor mortaliteit (R^2 = 0,44) en mRS (R^2 = 0,42), had het lineaire regressiemodel voor de EQ-5D de laagste goodness-of-fit (R^2 = 0,37). De grootste toename in R^2 van de modellen voor mortaliteit en mRS werd waargenomen na toevoeging van NIHSS, terwijl dit voor de EQ-5D index score was na toevoeging van leeftijd als voorspeller. Nadat de mRS en EQ-5D index scores beide als uitkomsten waren gedichotomiseerd, werden de AUC's vergeleken: AUC = 0,87 (mortaliteit) vs. AUC = 0,83 (mRS ≥ 3) vs. AUC = 0,78 (EQ-5D index score ≥ 0,65). In tegenstelling tot de modellen voor mortaliteit en mRS waren er meer voorspellers nodig in het model voor de EQ-5D index score om het voorspellend vermogen een plateau te bereiken.

Hoofdstuk 4 bouwt voort op de analyses van hoofdstuk 9, met het doel om case-mix correctiemodellen te bouwen voor individuele EQ-5D domein scores na een acute ischemische beroerte en om ziekenhuiseffect schattingen voor deze scores voor en na case-mix correctie te vergelijken. Een secundaire analyse werd uitgevoerd op een data subset (N = 336) uit de MR CLEAN trial, nadat multiple imputation (10 iteraties) werd uitgevoerd om missing kandidaat voorspellers te vervangen, patiënten werden geëxcludeerd die stierven vóór de drie maanden follow-up of die missing EQ-5D data hadden, en vier ziekenhuizen werden geëxcludeerd die minder dan 5 patiënten vertegenwoordigden. In deze substudie werden de punten op elk EQ-5D-domein gebruikt als uitkomst voor vijf individuele ordinale regressiemodellen. Veel voorkomende voorspellers in de regressiemodellen over alle individuele EQ-5D domeinen waren hypertensie en perifeer vaatlijden. R^2-waarden varieerden van 0,31 ("zelfzorg"), 0,27 ("mobiliteit"), 0,19 ("dagelijkse activiteiten"), 0,13 ("pijn/ongemak"), tot 0,08 ("angst/depressie"). Na case-mix correctie veran-

derde het geschatte effect van individuele beroerte centra slechts in geringe mate voor de “fysieke” domeinen in tegenstelling tot de “psychologische” domeinen.

DEEL II – PRAKTISCHE TOEPASSINGEN VAN PROMS IN DE KLINISCHE ZORG

Het tweede deel van dit proefschrift onderzocht niet alleen de implementatie van PROMs in de klinische zorg, inclusief de facilitators en barrières die worden waargenomen of ervaren door zorgprofessionals, maar ook de prospectief verzamelde PROMs van zowel gezonde vrouwen als borstkankerpatiënten.

De literatuurstudie in **Hoofdstuk 5** gaf een overzicht van 34 artikelen over PROM interventies (inclusief methoden en frequenties van verzamelen) in de zorg voor borstkanker, de facilitators en barrières ervan, en de impact ervan op patiënten, clinici en zorgprocessen. De meest gebruikte instrumenten in artikelen waren de CTCAE, de EORTC-QLQ-C30 en de EQ-5D. In de meeste geïncludeerde studies vond de verzameling van PROMs plaats tijdens de therapie, waarbij PROMs meestal elektronisch werden verzameld. Een veelbelovend effect van PROM collectie op patiënten (bijv. symptoomcontrole, kwaliteit van leven, en patiënttevredenheid), zorgverleners (bijv. compliance, monitoring van symptomen) en zorgprocessen (bijv. communicatie tussen arts en patiënt, ziekenhuisbezoeken) werd geobserveerd. De meest frequent gerapporteerde facilitators en barrières waren meestal van technische aard (bijv. IT-infrastructuur).

In **Hoofdstuk 6** werden analyses uitgevoerd van PROM scores van borstkankerpatiënten die een borstsparende operatie (N = 143) of een ablatio mamma (N = 70) ondergingen in het Erasmus MC. Binnen beide chirurgische strategieën werden PROMs (BREAST-Q, EORTC-QLQ-C30 en EORTC-QLQ-BR23) verzameld en vergeleken vanaf baseline (voorafgaand aan de operatie) tot 12 maanden na de primaire operatie. Multiple imputation werd gebruikt om missing PROM scores op T0, T6 en / of T12 te vervangen. Na 12 maanden follow-up was er een significante verbetering in “Body Image” (EORTC-QLQ-BR23) voor patiënten die een borstsparende operatie ondergingen, terwijl “Physical Wellbeing” (BREAST-Q), “Physical Functioning” (EORTC-QLQ-C30)), “Sexual Wellbeing” (BORST-Q) juist significant afnamen. Een statistisch significante afname van zowel “Breast Satisfaction” (BREAST-Q) als “Psychosocial Wellbeing” (BREAST-Q) scores werd waargenomen bij patiënten die een ablatio mamma ondergingen. Een matig sterke correlatie ($r = -0,75$, $p < 0,001$) werd waargenomen tussen “Physical Wellbeing” en “Breast Symptoms”

(EORTC-QLQ-BR23) op T12 voor patiënten na een borstsparende operatie, terwijl "Satisfaction with Breasts" en "Body Image" matig correleerden (r = 0,68, p <0,001) op T12 na een ablatio mamma.

Hoofdstuk 7 beschrijft een cross-sectionele studie waarin normatieve BREAST-Q scores uit een Nederlandse steekproef werden verzameld, in een poging om meer context te bieden voor zowel toekomstige patiënten als clinici bij het interpreteren van de scores van borstkankerpatiënten. Vrouwen werden uitgenodigd om zelf een digitale enquête in te vullen, inclusief demografische vragen en een Nederlandse versie van de BREAST-Q. In totaal werden vragenlijsten van 9059 vrouwen, zonder (voorgeschiedenis van) borstkanker, geanalyseerd. De gemiddelde normatieve scores voor de BREAST-Q domeinen waren 64.24 ± SD 18.60 ("Satisfaction with Breasts"), 71.95 ± SD 15.93 ("Psychosocial Wellbeing"), 89.54 ± SD 12.48 ("Physical Wellbeing") en 60.38 ± SD 15.37 ("Sexual Wellbeing"). Een borstoperatie in de voorgeschiedenis was geassocieerd met hogere "Satisfaction with Breasts" en "Sexual Wellbeing" scores, maar lagere "Physical Wellbeing" scores.

De cross-sectionele studie in **Hoofdstuk 8** naar de ervaring en perspectieven van zorgprofessionals over de implementatie van PROMs werd aangewakkerd door een verdere interesse in de facilitators en barrières nadat de literatuurstudie, beschreven in hoofdstuk 5, werd uitgevoerd. Een vragenlijst werd samengesteld, die zowel stellingen (met een 5-punts Likert schaal) als 8 open vragen bevat over facilitators en barrières van de implementatie van PROMs in de klinische zorg. 61 deelnemers (responspercentage 33%) in verschillende fasen van het implementatieproces werden geïncludeerd, voornamelijk medisch specialisten (51%) met een mediane arbeidsduur van 14 jaar. De meeste deelnemers waren betrokken bij de patiëntenzorg voor het syndroom van Turner, schisis en gynaecologische maligniteiten. Facilitators waren onder meer de aanwezigheid van een coördinator in de kliniek en de multidisciplinaire samenwerking. Frequent gerapporteerde barrières waren de uitdagende IT-infrastructuur (inclusief toegang tot het PROM platform, dashboard visualisatie, integratie van PROMs in het elektronisch patiëntendossier) en tijdsdruk in combinatie met de arbeidsintensieve aard van het incorporeren van PROMs in de zorg.

DISCUSSIE

Het doel van dit proefschrift was zowel de exploratie van de methodologie van het ontwikkelen van case-mix correctiemodellen voor zowel klinische als patiënt-

gerapporteerde uitkomsten bij CVA-patiënten, als de praktische toepassing (inclusief implementatie) van PROMs in de klinische zorg.

In de eerste helft van dit proefschrift werden case-mix correctiemodellen voor een generieke patiënt-gerapporteerde uitkomstmaat (EQ-5D) ontwikkeld ter vergelijking met modellen voor klinische uitkomsten en ter illustratie van belangrijke kanttekeningen bij de benchmarking van PROMs. Het voorspellend vermogen van de modellen voor EQ-5D was aanzienlijk minder in vergelijking met die van de modellen voor klinische uitkomsten, waarbij ook meer predictoren nodig waren in het EQ-5D model voordat het voorspellend vermogen een plateau bereikte. De verklaring kan worden gevonden in het ontbreken van aanvullende niet-traditionele baseline variabelen (bijv. psychosociale factoren), die potentieel meer bijdragen aan het voorspellend vermogen van modellen voor 'subjectieve' uitkomsten zoals PROMs. Ook werd aangetoond dat het voorspellend vermogen van modellen voor individuele (psychologische) EQ-5D-domeinen 'angst/depressie' en 'pijn/ongemak' lager was dan voor de (fysieke) domeinen 'dagelijks activiteiten', 'mobiliteit' en ' zelfzorg'. Deze bevindingen suggereren ook dat PROMs meer worden beïnvloed door een breed scala aan case-mix factoren dan traditionele (klinische) uitkomsten. Daarom, als het de wens is om PROMs te gebruiken voor benchmarking, dan zal het verzamelen van niet-traditionele case-mix factoren nodig zijn. Een belangrijk onderzoeksdoel in de toekomst is de identificatie van case-mix factoren die sterk geassocieerd zijn met frequent gebruikte PROMs en het effect van die factoren op het verschil in het voorspellend vermogen van case-mix correctiemodellen tussen klinische uitkomsten en PROMs.

De tweede helft van dit proefschrift richtte zich op de praktische toepassing van PROMs in de klinische zorg, inclusief verzamelmethoden, de analyse van prospectief verzamelde PROMs, het vergaren van normatieve PROM data, en de identificatie van facilitators en barrières van de implementatie van PROMs in de klinische zorg. PROMs hebben de potentie om de kwaliteit van zorg te verbeteren door een scherpere focus op waarde van zorg voor de patiënt en wat die belangrijk vindt. De longitudinale collectie van PROMs kan echter belemmerd worden wegens uitdagingen omtrent IT-infrastructuur, compliance van de patiënt in de loop van de tijd, en de tijdsdruk van zorgverleners. De toevoeging van PROMs in de klinische zorg heeft meer kans van slagen met een multidisciplinaire aanpak, waarbij artsen, verpleegkundigen, patiënten, management, onderzoekers, een data beheerder, IT- en administratief personeel betrokken zijn. Om de geïndividualiseerde zorg te bevorderen, moeten patiënten door zorgverleners worden geïnformeerd over het nut en belang van PROMs. Hierbij kan normatieve data een belangrijke rol spelen

om de PROM scores van patiënten te contextualiseren en de communicatie tussen patiënt en zorgverlener te verbeteren. Door continuerend onderzoek en het delen van expertise door zorgprofessionals, zal de toepassing van PROMs in de klinische zorg zich blijven evolueren.

Appendices

LIST OF SCIENTIFIC PUBLICATIONS

This Thesis:

- van Egdom LSE[†], **Oemrawsingh A**[†], Verweij LM, Lingsma HF, Koppert LB, Verhoef C, Klazinga NS, Hazelzet JA. Implementing Patient-Reported Outcome Measures in Clinical Breast Cancer Care: A Systematic Review. *Value Health. 2019 Oct;22(10):1197-1226. doi: 10.1016/j.jval.2019.04.1927.*

- **Oemrawsingh A**, van Leeuwen N, Venema E, Limburg M, de Leeuw FE, Wijffels MP, de Groot AJ, Hilkens PHE, Hazelzet JA, Dippel DWJ, Bakker CH, Voogdt-Pruis HR, Lingsma HF. Value-Based Healthcare in Ischemic Stroke Care: Case-Mix Adjustment Models for Clinical and Patient-Reported Outcomes. *BMC Med Res Methodol. 2019 Dec 5;19(1):229. doi: 10.1186/s12874-019-0864-z.*

- **Oemrawsingh A**, Swami N, Valderas JM, Hazelzet JA, Pusic AL, Gliklich RE, Bergmark RW. Patient-Reported Morbidity Instruments: A Systematic Review. *Value Health. 2020 Jun;23(6):791-811. doi: 10.1016/j.jval.2020.02.006.*

- **Oemrawsingh A**, van Leeuwen N, van der Lugt A, Klazinga NS, Hazelzet JA, van Busschbach JJ, Dippel DWJ, Lingsma HF; on behalf of MR CLEAN investigators. Case-mix models for hospital comparisons in EQ-5D domains after ischemic stroke: a substudy of the MR CLEAN trial. *Submitted.*

- **Oemrawsingh A**, van Leeuwen N, Mureau MAM, Lingsma HF, Verhoef C, Klazinga NS, Hazelzet JA, Koppert LB. A Prospective Analysis of Short-Term Patient-Reported Outcomes After Breast-Conserving Therapy and Mastectomy Without Reconstruction. *Submitted.*

- **Oemrawsingh A**[†], Clarijs ME[†], Pusic AL, Lingsma HF, Verhoef C, Hazelzet JA, Koppert LB. BREAST-Q Breast-Conserving Therapy Module: Normative Data From A Dutch Sample of 9059 Women. *Accepted for publication..*

- Amini M[†], **Oemrawsingh A**[†], Verweij LM, Klazinga NS, Lingsma HF, Hazelzet JA, Eijkenaar F, van Leeuwen N. Facilitators and Barriers of Implementing Patient-Reported Outcome Measures in Clinical Care: an Academic Center's Initial Experience. *Health Policy. 2021 Jul 8:S0168-8510(21)00173-1.*

Other Publications:

- **Oemrawsingh A**, de Boer NL, Brandt-Kerkhof ARM, Verhoef C, Burger JWA, Madsen EVE. Short-Term Complications in Elderly Patients Undergoing CRS and HIPEC: A Single Center's Initial Experience. *Eur J Surg Oncol. 2019 Mar;45(3):383-388. doi: 10.1016/j.ejso.2018.10.545.*

- van Kooten JP[ǂ], **Oemrawsingh A**[ǂ], de Boer NL, Verhoef C, Burger JWA, Madsen EVE, Brandt-Kerkhof ARM. Predictive Ability of C-Reactive Protein in Detecting Short-Term Complications After Cytoreductive Surgery and Hyperthermic Intraperitoneal Chemotherapy: A Retrospective Cross-Sectional Study. *Ann Surg Oncol. 2020 Jun 10. doi: 10.1245/s10434-020-08619-y.*

- **Oemrawsingh A**, Hazelzet JA, Koppert LB. "Hoofdstuk 17 Patiëntgerichte borstkankerzorg: ervaring en perspectief van een academisch centrum". NFU bundel *Gepersonaliseerde medische zorg – Innovatieve zorg afgestemd op persoonlijke behoeften, voorkeuren en waarden*. Juni, 2020. ISBN: 978-90-9033-183-6.

- Devarakonda KS, Timman R, Bouvy PF, **Oemrawsingh A**, Apon I, Mureau MAM, Koppert LB, Kranenburg LW. Trends in Emotional Functioning and Psychosocial Wellbeing in Breast Cancer Survivors. *Submitted.*

- Clarijs ME, **Oemrawsingh A**, Bröker MEE, Verhoef C, Lingsma HF, Koppert LB. Quality of Life of Caregivers of Breast Cancer Patients: a Cross-Sectional Evaluation. *Submitted.*

[ǂ] **= shared co-first authorship**

PATIËNTGERICHTE BORSTKANKERZORG: ERVARING EN PERSPECTIEF VAN EEN ACADEMISCH CENTRUM

Bijdragend hoofdstuk bij de uitgave van J.A. Hazelzet en N. van Weert (namens de Nederlandse Federatie van UMC's), getiteld "Gepersonaliseerde medische zorg – Innovatieve zorg afgestemd op persoonlijke behoeften, voorkeuren en waarden". Juni, 2020. ISBN: 978-90-9033-183-6

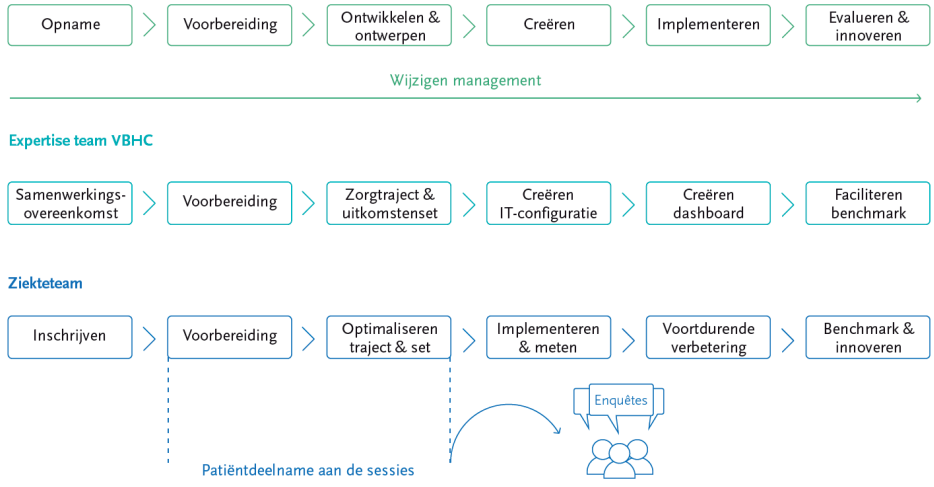
Arvind Oemrawsingh, Jan A. Hazelzet, Linetta B. Koppert

In Nederland wordt bij ongeveer één op de zeven vrouwen borstkanker vastgesteld. [1,2] In de praktijk gebruiken we steeds vaker patiënt-gerapporteerde uitkomsten (PRO's) om resultaten van de behandeling van borstkanker te meten.

Deze PRO's gaan bijvoorbeeld over kwaliteit van leven en het dagelijks functioneren van patiënten na een behandeling. Het belang van dergelijke uitkomsten groeit, nu de overlevings- en andere klinische uitkomsten voor borstkankerpatiënten (in het beginstadium van de ziekte) goed zijn, ongeacht welke borstoperatie is uitgevoerd.[3-5] PRO's laten zien wat het (neven) effect van de behandeling is op het dagelijks leven van vrouwen, zoals beweeglijkheid van de arm, pijn en cosmetisch resultaat. Zo vormen ze dé basis voor persoonsgerichte zorg op maat en voor het verbeteren van de borstkankerzorg.

Het Academisch Borstkankercentrum van het Erasmus MC behandelt veel vrouwen die op jonge leeftijd borstkanker krijgen, soms tijdens een zwangerschap. Vaker dan anders speelt bij deze groep patiënten erfelijkheid een rol. Ook maken deze nog jonge patiënten zich regelmatig zorgen over hun vruchtbaarheid na chemotherapie en staan zij voor complexe keuzes over verschillende vormen van borstoperaties, hersteloperaties en bestraling.

Deze ervaring was de aanleiding om een multidisciplinair zorgpad te ontwikkelen met ruimte om op verschillende momenten uitkomsten te meten, ook patiëntgerapporteerde uitkomsten. Dit zorgpad vormde de basis voor de waardegedreven borstkankerzorg in het Erasmus MC[6] (figuur 1). Het algemene doel was om bereikte uitkomsten te meten in plaats van het volume van de geleverde diensten.[7]



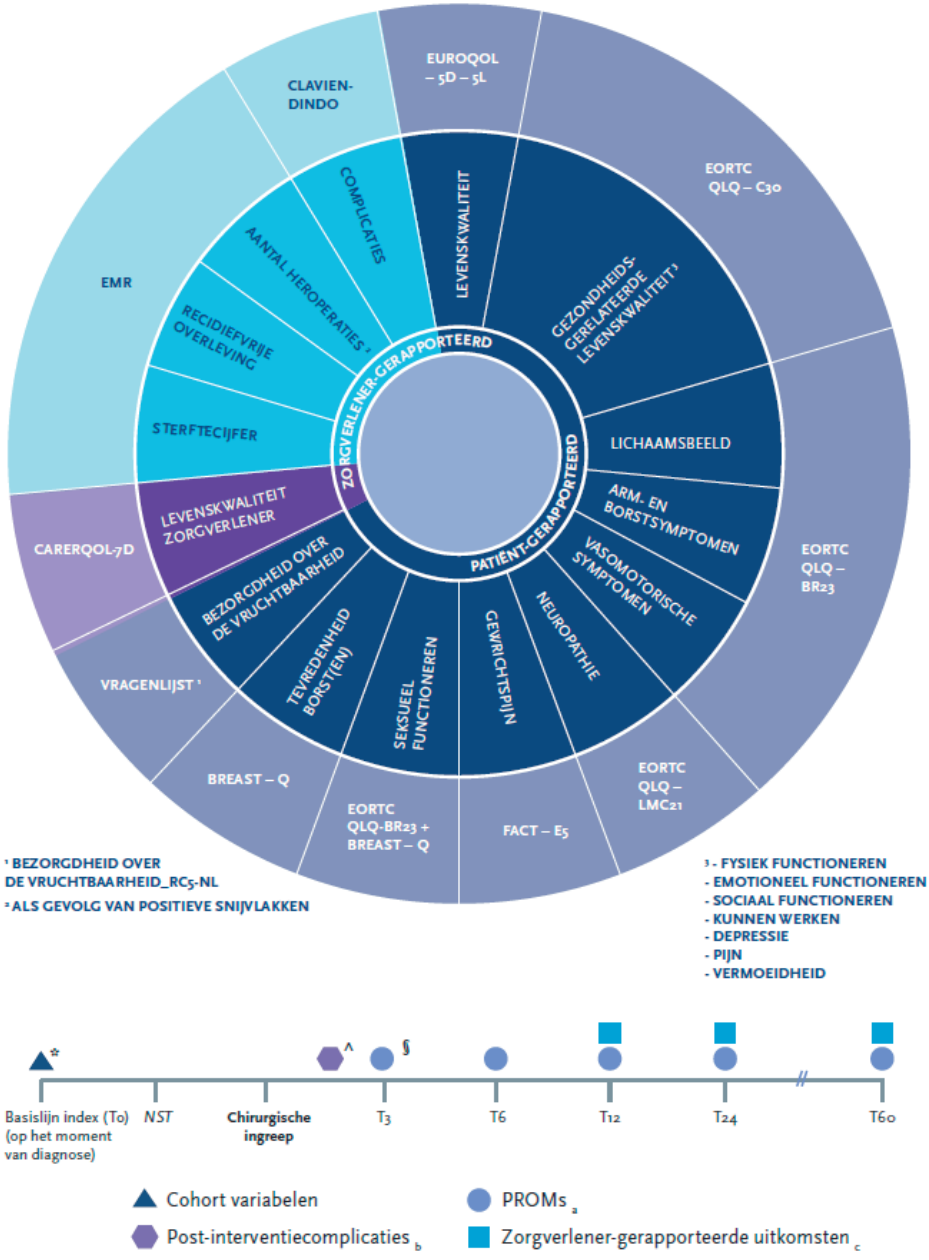
Figuur 1 | De blauwdruk van het Erasmus MC ondersteunt de teams in hun ontwikkeling naar waardegedreven zorg.[6]

Onze keuze voor waardegedreven zorg sloot aan bij de keuze van de Raad van Bestuur van het Erasmus MC om deze academische zorginstelling te hervormen naar een op waarde gebaseerd kader voor de zorg. Het multidisciplinaire borstkankerteam (waaronder medisch oncologen, chirurgisch oncologen, plastische en reconstructieve chirurgen, radiotherapeut-oncologen, borstkankerverpleegkundigen, klinische genetici en psychologen) greep deze gelegenheid aan om de zorg voor borstkankerpatiënten stelselmatig te meten en zo beter te laten aansluiten bij de behoefte van de individuele patiënt.

Het team startte met het herontwerpen van het zorgtraject (zorgpad), waarin momenten voor vervolgconsulten binnen verschillende disciplines én evaluatiemomenten en uitkomstmetingen werden bepaald. Dit zijn zowel klinische (bijvoorbeeld sterftecijfers, complicatiecijfers, heropnamecijfers) als patiënt-gerapporteerde uitkomstmetingen (PROMs), waarvoor patiënten vragenlijsten invullen. De te gebruiken PROMs moesten voornamelijk ziektespecifiek zijn en een beeld geven van de gehele zorgcyclus, de ziektelast op de lange termijn en de kwaliteit van leven van de patiënt.

Ontwikkeling van patiënt-gerapporteerde uitkomstmaten
Het team wilde uitkomsten meten die patiënten zelf belangrijk vinden na een borstkankerbehandeling. Daarom werd voor de ontwikkeling van een ziektespecifieke uitkomstenset aan patiënten in verschillende stadia van de behandeling gevraagd welke uitkomsten zij het belangrijkst vonden. Na een uitgebreide zoektocht naar passende gevalideerde vragenlijsten kozen we een uitkomstenset die zowel generieke (bijvoorbeeld EQ-5D-5L, EORTC-QLQ-C30)[8] als borstkankerspecifieke vragenlijsten (bijvoorbeeld BREAST-Q, EORTC-QLQ-BR23)[9] omvat (figuur 2). De meeste PROMs worden voorgelegd bij de diagnose (tijdstip 0), zes maanden na de primaire chirurgische behandeling (tijdstip 6), en vervolgens jaarlijks (tijdstip 12-60) (figuur 2). Naast deze PROMs legden we case-mixvariabelen en het behandelplan vast.

Tegelijkertijd heeft het Academisch Borstkankercentrum van het Erasmus MC in samenwerking met het (ICHOM) bijgedragen aan de ontwikkeling van een internationale gestandaardiseerde uitkomstenset voor borstkanker, die zowel klinische als patiënt-gerapporteerde uitkomsten omvat.[10] De vastgestelde (vervolg)tijdstippen voor uitkomstmetingen in de ICHOM-borstkankerset zijn vergelijkbaar met de tijdstippen in ons zorgpad.



Figuur 2 | Erasmus MC standaardset voor borstkanker.[6]

* Preoperatieve PROMs

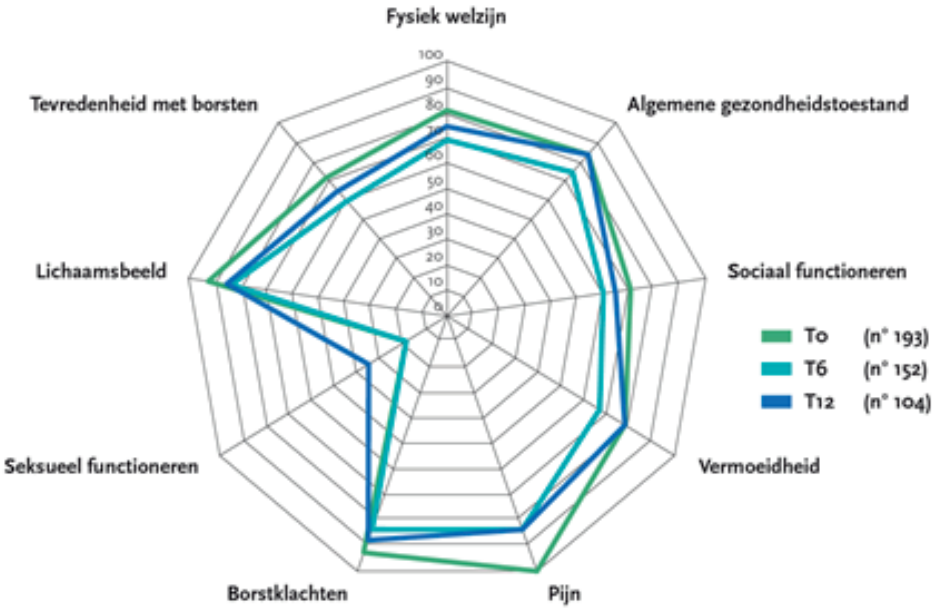
^ Door de zorgverlener gemelde complicaties gemeten binnen 90 dagen na de operatie

§ Enkel in geval van neo-adjuvante systeemtherapie (NST)

Instrument voor dataverzameling, web-based platform met dashboard

De ontwikkeling van een instrument voor dataverzameling en een veilig platform was een volgende stap op weg naar implementatie van waardegedreven zorg.[6] Ons elektronische instrument voor dataverzameling maakt, na selectie van het juiste zorgtraject, automatische verzending van PROMs naar patiënten mogelijk. De vragenlijsten worden automatisch via e-mail verstuurd op de vastgestelde tijdstippen, voorafgaand aan het geplande consult. Als de PROMs nog niet zijn ingevuld, volgt tot drie weken voor het consult wekelijks een herinnering per e-mail.

Met twee softwareprogramma's, LimeSurvey en GemsTracker, werd een veilig *web-based* platform gebouwd. Door de koppeling van het instrument voor dataverzameling aan het elektronisch patiëntendossier en de ontwikkeling van een gebruiksvriendelijk dashboard, kunnen artsen het verloop van de PROM-scores vóór de consulten in de polikliniek beoordelen (figuur 3). Artsen melden een verbeterd begrip van de klachten of problemen in kwaliteit van leven van de individuele patiënt. Dit heeft artsen ook aangemoedigd om gevoelige onderwerpen (bijvoorbeeld seksualiteit) aan te kaarten, die anders onbesproken zouden zijn gebleven.



Figuur 3 | Spindiagram dat de verschillende PROM-scores per tijdstip weergeeft (239 patiënten).[6]

Regionale oncologienetwerken
Om gezamenlijk de kwaliteit van de borstkankerzorg te waarborgen en te verbeteren, tekenden acht borstkankercentra uit Zuidwest-Nederland in 2018 een overeenkomst voor de implementatie van waardegedreven borstkankerzorg. Elk deelnemend ziekenhuis is zelf voor de implementatie verantwoordelijk. Een onderdeel is het systematisch verzamelen van PRO-gegevens binnen de regio voor vergelijkingsdoeleinden (benchmarking). De resultaten kunnen zo leiden tot een verbetering van de kwaliteit van de zorg in de regio. Ook werd gezamenlijk besloten dat alle deelnemende borstkankercentra de vastgestelde ICHOM-borstkanker-uitkomstenset gaan gebruiken. Dit vergemakkelijkt in de toekomst ook internationale benchmarking.

Dutch Hospital Data (DHD), opgericht door de Nederlandse Vereniging van Ziekenhuizen (NVZ) en de Nederlandse Federatie van Universitair Medische Centra (NFU), kan ziekenhuizen helpen bij het uniform vastleggen en integreren van klinische en patiënt-gerapporteerde uitkomsten in hun elektronisch patiëntendossier. DHD beheert en verwerkt diverse gegevensbestanden van ziekenhuizen en ondersteunt de levering van data, die voldoen aan strenge privacy voorwaarden.

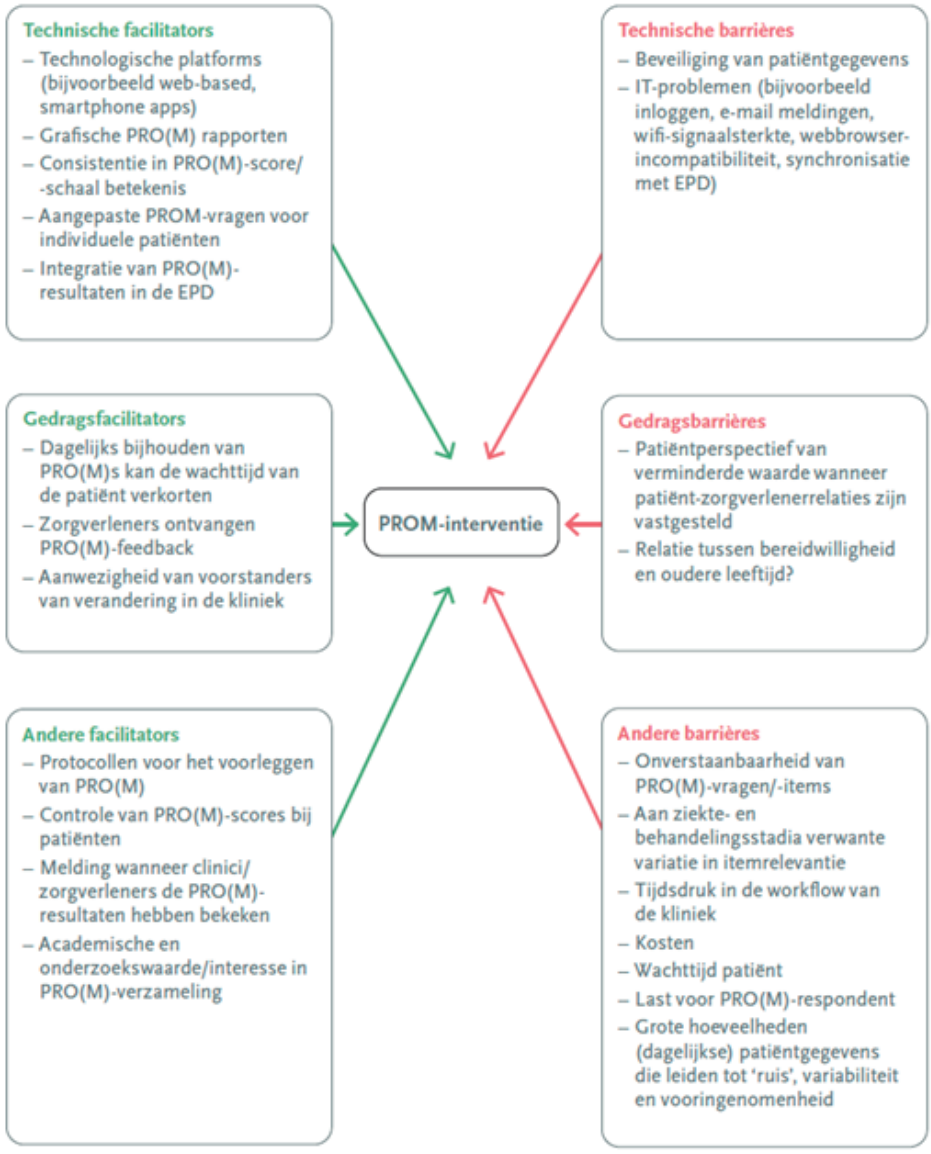
Voorgaande onderzoeksprojecten als basis
Het Academisch Borstkankercentrum van het Erasmus MC deed al eerder onderzoek naar de mogelijkheden van PROMs. In samenwerking met Borstkankervereniging Nederland (BVN) werd onderzoek uitgevoerd om de verschillen tussen de geschatte PROM-scores en de normatieve en referentiescores aan te tonen. Ook zijn de tevredenheid en verwachtingen van borstkankeroverlevenden aan de hand van PROMs geëvalueerd. Op basis van de antwoorden van bijna vijfhonderd patiënten (die meer dan zes maanden geleden geopereerd waren) vonden we dat de geschatte EORTC-QLQ-C30 en EORTC-QLQ-BR23 scores niet significant verschilden van de normatieve en referentiescores. De enige uitzondering was het domein 'fysiek welbevinden' (BREAST-Q), waarbij in dit onderzoek lagere scores werden gerapporteerd in vergelijking met normatieve scores. De vergelijking van PROM-scores met normatieve en referentiescores zou een adequate evaluatie van de kwaliteit van de geleverde zorg mogelijk kunnen maken. De auteurs concluderen ook dat de meerderheid van de patiënten het voorleggen van PROMs 'aanvaardbaar' of 'zeer aanvaardbaar' vond. Patiënten meldden dat de uitkomsten kunnen worden gebruikt als een eigen 'begeleidingsinstrument' tijdens de individuele zorg. Maar ze gaven ook aan dat routinematig gebruik van PROMs de kwaliteit van de zorg in het algemeen kan verbeteren.

Een samenwerking tussen de afdelingen Chirurgische Oncologie en Medische Besliskunde in het Erasmus MC leidde in 2018 tot de publicatie van een systematische review van methoden voor het verzamelen van PROMs waarin zowel de faciliterende aspecten (facilitators) als barrières in de klinische borstkankerzorg[12] werden belicht. In het gros van de 34 artikelen werd gemeld dat het verzamelen van PROMs een veelbelovende positieve invloed had op patiënten: therapietrouw, minder klachten, hoge aanvaardbaarheid en tevredenheid over de verzamelmethode.

Zorgverleners noemden een beter onderbouwde klinische besluitvorming en betere symptoombestrijding als positieve punten. Op het gebied van zorgprocessen/-structuren noemden ze als voordelen: meer doorverwijzingen, minder bezoeken aan de spoedeisende hulp en betere communicatie tussen patiënt en zorgverlener. Vaak beschreven succesfactoren (facilitators, figuur 4) waren het gebruik van technologische verzamelmethoden (via tablet of smartphone) en een klinische leider als voortrekker. Vaak beschreven barrières in de studies waren technische problemen (zoals het gebrek aan integratie van een PRO-dashboard in het elektronisch patiëntendossier, problemen met het inloggen op het platform, gegevensbeveiliging) en tijdsbeperkingen in de workflow van de polikliniek.

In 2018 werd nog een ander onderzoek gepubliceerd door het Academisch Borstkankercentrum met als doel referentie-PROM-scores te verkrijgen en associaties te bepalen tussen de leeftijd van de patiënten, de tijd sinds de borstoperatie, het type borstoperatie en andere voorspellers van PROM-scores.[13]

Met terugwerkende kracht werden PROMs (EQ-5D, EORTC-QLQ-C30/ BR23 en BREAST-Q) geanalyseerd van ongeveer 760 pTis-T3N0-3M0 borstkankerpatiënten die tussen 2005 en 2016 een borstkankeroperatie hadden ondergaan. Eén van de waarnemingen was dat de meeste PRO-modules een hoog voltooiingspercentage hadden (88-100%), met uitzondering van ‘seksueel welbevinden’ (BREAST-Q). Multivariabele regressie toonde aan dat alleen mastectomie (borstamputatie) en reconstructies met implantaten werden geassocieerd met een lagere score voor ‘tevredenheid met borsten’ (BREAST-Q) in vergelijking met borstsparende operaties (BCT). Aanzienlijk lagere scores hiervoor werden ook geassocieerd met bestralingstherapie. Toenemende leeftijd werd geassocieerd met afnemende scores voor ‘fysiek welbevinden’ en ‘seksueel welbevinden’ (EORTC-QLQ-C30).



Figuur 4 | Beschrijving van facilitators en barrières.[12]

Huidige en toekomstige projecten

De afgelopen vier jaar heeft het Academisch Borstkankercentrum een aanzienlijke hoeveelheid patiëntgebonden uitkomstendata verzameld. Deze PROMs werden prospectief verzameld als onderdeel van de waardegedreven borstkankerzorg. Onderzoekers van de afdeling Medische Besliskunde en Chirurgische Oncologie analyseren samen deze gegevens om meer kennis te vergaren over algemene

en borstkankerspecifieke kwaliteit-van-leven aspecten binnen deze populatie. Daarnaast wordt onderzoek gedaan naar methodologische kwesties (bijvoorbeeld statistische onzekerheid, aanpassing van de cohort en het omgaan met ontbrekende (PRO-)data, hetgeen van invloed kan zijn op de analyse en interpretatie van PROMs).

Het doel is om deze gegevens in de polikliniek te gebruiken bij het patiëntconsult voorafgaand aan de operatie én tijdens de follow-up. Door patiënten vóór de operatie het verloop van de PROM-scores van eerder behandelde patiënten te tonen, kan dit meer context bieden voor de keuze tussen verschillende (chirurgische) behandelmogelijkheden. Dit kan leiden tot een beter gezamenlijk besluitvormingsproces en minder spijt achteraf over de beslissing.

Eén onderzoek toonde bijvoorbeeld het verloop aan van de PROM-scores binnen verschillende chirurgische behandelstrategieën gedurende een periode van een jaar (borstsparende operatie ten opzichte van borstamputatie alleen versus borstamputatie plus borstreconstructie).[15] Een analyse van bijna 260 borstkankerpatiënten toont een significante vermindering aan van de mediaanscores voor 'tevredenheid met borsten' gedurende twaalf maanden voor patiënten die alleen een borstamputatie of een borstamputatie plus een reconstructieve borstoperatie ondergingen. Ook werd een significante verlaging van de scores voor 'seksueel welbevinden' gedurende een jaar waargenomen bij patiënten die een borstsparende operatie ondergingen. Het verloop van de PROM scores tijdens de follow-up kan zowel patiënten als clinici helpen om de impact van de chirurgische behandeling van borstkanker beter te begrijpen en de verwachtingen te managen.

Komende onderzoeken richten zich op het vaststellen van Nederlandse normatieve scores en de scores voor het minimaal klinisch relevante verschil (Minimal Clinically Important Difference, MCID) voor het BREAST-Q instrument.[16] Een MCID-score geeft de kleinste verandering in de PROM-score aan die door patiënten als belangrijk wordt ervaren.[17] Deze onderwerpen vragen om vervolgonderzoek, omdat dit de interpreteerbaarheid en daarmee het nut van PROMs in de patiëntgerichte zorg en in vergelijkend effectiviteitsonderzoek zal vergroten. In het kader van het NFU-programma Waardegedreven Zorg vormen vier universitair medische centra een samenwerkingsverband in de borstkankerzorg, met de afspraak om dezelfde uitkomstenset te gebruiken.

Als onderdeel van de European University Hospital Alliance (EUHA) heeft het Erasmus MC zich samen met acht andere universitaire ziekenhuizen toegelegd op

het stimuleren van innovatie, het delen van expertise op het gebied van gezondheidszorg en onderzoek en het leren van elkaar, met als doel de kwaliteit van de zorg te optimaliseren. Daarom zijn het Academisch Borstkankercentrum en de andere internationale partners van plan om klinische en patiënt-gerapporteerde data te delen ten behoeve van benchmarking. De meeste universitaire ziekenhuizen maken gebruik van de ICHOM borstkanker-uitkomstenset.

REFERENTIES

1. https://www.rivm.nl/bevolkingsonderzoek-borstkanker.
2. https://www.cijfersoverkanker.nl/.
3. Lagendijk M, van Maaren MC, Saadatmand S, Strobbe LJA, Poortmans PMP, Koppert LB, et al. Breast conserving therapy and mastectomy revisited: Breast cancer-specific survival and the influence of prognostic factors in 129,692 patients. *Int J Cancer*. 2018;142(1):165-75.
4. Litiere S, Werutsky G, Fentiman IS, Rutgers E, Christiaens MR, Van Limbergen E, et al. Breast conserving therapy versus mastectomy for stage I-II breast cancer: 20 year follow-up of the EORTC 10801 phase 3 randomised trial. *Lancet Oncol*. 2012;13(4):412-9.
5. Poggi MM, Danforth DN, Sciuto LC, Smith SL, Steinberg SM, Liewehr DJ, et al. Eighteen-year results in the treatment of early breast carcinoma with mastectomy versus breast conservation therapy: the National Cancer Institute Randomized Trial. *Cancer*. 2003;98(4):697-702.
6. Van Egdom LSE, Lagendijk M, van der Kemp MH, van Dam JH, Mureau MAM, Hazelzet JA, et al. Implementation of Value Based Breast Cancer Care. *Eur J Surg Oncol*. 2019;45(7):1163-70.
7. Porter MTE, Teisberg Olmsted E. Redefining Health Care: Creating Value-Based Competition on Results. Boston: Harvard Business School Press. 2006.
8. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *JNCI: Journal of the National Cancer Institute*. 1993;85(5):36576.
9. Pusic AL, Klassen AF, Scott AM, Klok JA, Cordeiro PG, Cano SJ. Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. *Plast Reconstr Surg*. 2009;124(2):345-53.
10. Ong WL, Schouwenburg MG, van Bommel ACM, Stowell C, Allison KH, Benn KE, et al. A Standard Set of Value-Based Patient-Centered Outcomes for Breast Cancer: The International Consortium for Health Outcomes Measurement (ICHOM) Initiative. *JAMA Oncology*. 2017;3(5):677-85.
11. Lagendijk M, van Egdom LSE, Richel C, van Leeuwen N, Verhoef C, Lingsma HF, et al. Patient reported outcome measures in breast cancer patients. *Eur J Surg Oncol*. 2018;44(7):963-8.
12. Van Egdom LSE, Oemrawsingh A, Verweij LM, Lingsma HF, Koppert LB, Verhoef C, et al. Implementing Patient-Reported Outcome Measures in Clinical Breast Cancer Care: A Systematic Review. *Value Health*. 2019;22(10):1197-226.
13. Lagendijk M, van Egdom LSE, van Veen FEE, Vos EL, Mureau MAM, van Leeuwen N, et al. Patient-Reported Outcome Measures May Add Value in Breast Cancer Surgery. *Annals of Surgical Oncology*. 2018;25(12):3563-71.
14. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Statistical Methods in Medical Research*. 2014;23(5):440-59.
15. Oemrawsingh A, van Leeuwen N, Mureau MAM, Lingsma HF, Verhoef C, Klazinga NS, Hazelzet JA, Koppert LB. A Prospective Analysis of Longitudinal Patient-Reported Outcomes Within Breast Cancer Surgical Treatment Strategies (Submitted). 2020.
16. Cano SJ, Klassen AF, Scott A, Alderman A, Pusic AL. Interpreting clinical differences in BREAST-Q scores: minimal important difference. *Plast Reconstr Surg*. 2014;134(1):173e-5e.
17. Cook CE. Clinimetrics Corner: The Minimal Clinically Important Change Score (MCID): A Necessary Pretense. *J Man Manip Ther*. 2008;16(4):E82-E3.

CONTRIBUTING AUTHORS

Aad van der Lugt
Department of Radiology & Nuclear Medicine
Erasmus University Medical Center

Aafke J. de Groot
Department of General Practice and Elderly Care Medicine
Amsterdam University Medical Center
Vivium Naarderheem

Andrea L. Pusic
Division of Plastic and Reconstructive Surgery Patient Reported Outcomes, Value and Experience (PROVE) Center
Brigham and Women's Hospital

Carla H. Bakker
Expert Centre Quality Registries
Leiden University Medical Center

Cornelis Verhoef
Academic Breast Cancer Center
Erasmus MC Cancer Institute
Erasmus University Medical Center

Diederik W.J. Dippel
Department of Neurology
Erasmus University Medical Center

Esmee Venema
Center for Medical Decision Making
Department of Public Health
Department of Neurology
Erasmus University Medical Center

Frank Eijkenaar
Erasmus School of Health Policy and Management
Erasmus University Rotterdam

Frank-Erik de Leeuw
Department of Neurology
Donders Institute for Brain Cognition and Behaviour
Radboud University Medical Center

Helene R. Voogdt – Pruis
EnCorps
Stroke Knowledge Network Netherlands

Hester F. Lingsma
Center for Medical Decision Making
Department of Public Health
Erasmus University Medical Center

Jan A. Hazelzet
Department of Public Health
Erasmus University Medical Center

Jan J. van Busschbach
Section of Medical Psychology & Psychotherapy
Department of Psychiatry
Erasmus University Medical Center

José M. Valderas
International Society for Quality of Life Research (ISOQOL)
Health Services & Policy Research
University of Exeter Medical School

Laurentine S.E. van Egdom
Academic Breast Cancer Center
Erasmus University Medical Center

Linetta B. Koppert
Academic Breast Cancer Center
Erasmus University Medical Center

Lisanne M. de Jong – Verweij
Department of Public Health
Erasmus University Medical Center

Marc A.M. Mureau
Department of Plastic and Reconstructive Surgery Erasmus MC Cancer Institute
Erasmus University Medical Center

Markus P. Wijffels
Department of Neurorehabilitation
Rijndam Rehabilitation

Marloes E. Clarijs
Academic Breast Cancer Center
Erasmus University Medical Center

Martien Limburg
Department of Neurology
Flevoziekenhuis
Stroke Knowledge Network Netherlands

Niek S. Klazinga
Department of Public Health
Academic Medical Centre
University of Amsterdam
Amsterdam Public Health Research Institute

Nikki van Leeuwen
Center for Medical Decision Making
Department of Public Health
Erasmus University Medical Center

Nishwant Swami
University of Massachusetts Medical School

Pieter H.E. Hilkens
Department of Neurology
St. Antonius Hospital

Regan W. Bergmark
Center for Surgery and Public Health
Patient Reported Outcomes, Value and Experience (PROVE) Center
Brigham and Women's Hospital
Department of Otolaryngology – Head and Neck Surgery
Harvard Medical School

Richard E. Gliklich
Department of Otolaryngology – Head and Neck Surgery
Harvard Medical School

PRESS CLIPPINGS

Systematic Collection of Patient-Reported Outcome Measures in Breast Cancer Care Has a Promising Impact on Patients, Providers, and Care Processes

Published Oct 21, 2019

Lawrenceville, NJ, USA—October 21, 2019—***Value in Health***, the official journal of ISPOR—the professional society for health economics and outcomes research, announced today the publication of new research showing that systematic collection of patient-reported outcome measures in breast cancer care has a promising impact on patients, providers, and care processes/systems. The report, "Implementing Patient-Reported Outcome Measures in Clinical Breast Cancer Care: A Systematic Review," was published in the October 2019 issue of ***Value in Health***.

In this systematic literature review, researchers screened a total of 2311 published articles, of which 34 eligible articles were ultimately included. The majority of studies described a promising effect of patient-reported outcome measures collection on patients (adherence, symptom distress, quality of life, acceptability, and satisfaction), providers (willingness to comply, clinical decision making, symptom management), and care process or system outcomes (referrals, patient-provider communication, hospital visits). A limited number of facilitators and barriers were identified, primarily of a technical and behavioral nature.

"While patient-reported outcome measures (PROMs) are increasingly being collected and advocated in cancer care, reviews focusing on methods of PROM administration, specifically in breast cancer care, have not been previously published," said author Jan A. Hazelzet, MD, PhD, Department of Public Health, Erasmus MC, Rotterdam, The Netherlands. "Our review found that systematic PROM collection in routine breast cancer care has a promising impact on patients, providers, and care processes/systems. If routine PROM collection is to be implemented in clinical practice, we recommend a standardized PROM set because it is one of the requirements for benchmarking treatments and healthcare providers. It could also ultimately allow comparisons of breast cancer outcomes across countries. By doing so, a major step will have been taken toward value-based healthcare."

Breast cancer is the most common cancer affecting women worldwide. "With survival rates for early-stage breast cancer patients continuing to improve, more attention must be paid to quality of life issues reported by patients themselves," added coauthor and Erasmus colleague, Arvind Oemrawsingh, MD, MHS. There is increasingly more focus being put on patient-reported outcome measures, as part of healthcare's shift towards a more value-based framework for quality of care improvement. Patient-reported outcomes are defined as feedback on a patient's health condition (ie, symptoms and quality of life) coming directly from the individual patient, thus without external interpretation. Measurement of these outcomes is based on self-completed questionnaires called patient-reported outcome measures, which are increasingly being collected and advocated in cancer care for aiding care management of the individual patient.

'Vergelijken van PROMs vergt andere correctie'

Datum: 9 jan '20 | **Soort:** Nieuws | **Auteur:** Maaike Zweers |

PROM-scores vergelijken tussen groepen of tussen ziekenhuizen vergt een andere correctie dan die van klinische uitkomstmaten. Er is bijvoorbeeld een correctie nodig op basis van sociale patiëntkenmerken. Dit blijkt een studie onder CVA-patiënten.

De studie is uitgevoerd in het kader van het programma Waardegedreven Zorg van het NFU-consortium Kwaliteit van Zorg. Voor een goede vergelijking zullen dus niet-klinische achtergrondkenmerken aan de data van CVA-patiënten gekoppeld moeten worden. De auteurs sluiten in een aanvullende toelichting niet uit dat dit ook geldt voor andere ziektebeelden.

Onbekend terrein

In Nederland is er een groeiende belangstelling ontstaan voor het gebruik van PROMs als bron van transparante informatie over de kwaliteit van zorg. De standaardisatie van PROMs voor het vergelijken van ziekenhuizen is nog onbekend terrein. De studie van Arvind Oemrawsingh e.a. onderzoekt de methodologie van het ontwikkelen van case-mix correctiemodellen voor zowel klinische als patiënt-gerapporteerde uitkomsten bij CVA-patiënten.

Variatie

De belangrijkste bevinding is dat er verschillen zijn in case-mix factoren tussen de modellen. Socio-economische status lijkt bijvoorbeeld een belangrijke case-mix factor te zijn die bijdraagt aan de variatie in de EQ-5D scores. Dit instrument dient om de gezondheidstoestand te bepalen aan de hand van door de patiënt te beantwoorden vragen. Het lijkt niet van invloed op de klinische uitkomsten.

Zoektocht naar meetinstrument voor morbiditeit

10 September 2020

Patiëntkenmerken als leeftijd en het hebben van meerdere aandoeningen (morbiditeit) zijn van invloed op behandeluitkomsten zoals sterfte, complicaties, opnameduur, heropnames en kwaliteit van leven. Behalve medische dossiers en administratieve codes, kunnen mogelijk ook patiënt-gerapporteerde morbiditeitsinstrumenten gebruikt worden als databron. Maar welke vragenlijsten zijn er en zijn ze ook betrouwbaar en bruikbaar? Arvind Oemrawsingh, arts-onderzoeker Erasmus MC, deed hier uitgebreid literatuuronderzoek naar.

"We kunnen behandeluitkomsten pas vergelijken, wanneer er voor bepaalde karakteristieken van patiëntpopulaties, de zogenaamde case-mix, goed gecorrigeerd kan worden. Dat lukt nog onvoldoende voor morbiditeit, omdat ziekenhuizen dit op verschillende manieren meten", vertelt Oemrawsingh. Reden genoeg om tijdens zijn onderzoeksstage bij ICHOM in Boston overzicht te bieden in bestaande vragenlijsten hiervoor.
Bij dit internationale consortium voor uitkomstmetingen in de zorg was hij hiervoor op het juiste adres. "De missie van ICHOM is om wereldwijd te komen tot uniforme gestandaardiseerde uitkomstensets die belangrijk zijn voor de patiënt. Alleen zo kunnen we gericht kwaliteit vergelijken en verbeteren. Zo'n set bestaat uit case-mix-factoren, en klinische en patiënt-gerapporteerde uitkomstmaten (PROMs). Ons onderzoek gaat over vragenlijsten naar een belangrijke case-mix-factor."

Zoektocht
In de Nederlandse praktijk wordt morbiditeit meestal geregistreerd op basis van medische dossiers. Volgens Oemrawsingh heeft dit beperkingen en zijn de gegevens niet altijd actueel. Daarom is ook patiëntonderzoek nodig. "Patiënten zijn zich tegenwoordig ook bewuster van hun gezondheid en hebben kennis over hun ziektebeeld en de medicatie die ze gebruiken. Ik verwacht dat we met vragenlijsten beter in kaart kunnen brengen wat de ernst en ziektelast van een extra aandoening als diabetes voor de patiënt betekent."
In zijn zoektocht selecteerde hij 34 uit ruim duizend mogelijk relevante artikelen over algemene morbiditeitsinstrumenten. Vervolgens zoomde hij in op de drie meest geciteerde en gevalideerde morbiditeitsvragenlijsten: de Self-Administered Comorbidity Questionnaire (SCQ/ waarvan een Nederlandse vertaling is gevalideerd), Self-Reported Charlson Comorbidity Index (SR-CC) en de Disease Burden Morbidity Assesment (DBMA).

Bruikbaar
Maar deze meetinstrumenten hebben nog de nodige haken en ogen en moeten nog verder ontwikkeld worden, is de bevinding van Oemrawsingh. Zo is verdere ontwikkeling van de betrouwbaarheid nodig. "Sommige studies bepaalden in hoeverre de uitkomsten van de vragenlijsten klopten met het medisch dossier. Bij diabetes mellitus was de betrouwbaarheid hoog, maar bij ziektebeelden als maagzweren en artrose waren de verschillen tussen de databronnen nog te groot. De betrouwbaarheid/ validiteit behoeft dus nog verbetering."
Daarnaast blijkt uit zijn onderzoek dat sommige instrumenten mortaliteit konden voorspellen maar dat het verband met procesmaten als opnameduur, heropnames en zorgkosten nog te zwak is.

Ook voor het realiseren van waardegedreven zorg is het belangrijk om morbiditeit mee te nemen als case-mix, vindt Oemrawsingh. "Een van de doelen van waardegedreven zorg is het sturen op patiënt-gerapporteerde uitkomsten (PROMs). Dan wil je weten wat het verband is tussen morbiditeitsvragenlijstscores en PROMs, zodat je na case-mix-correctie kunt ontdekken waar je waarde kunt toevoegen."
Oemrawsingh hoopt dan ook dat zijn studie meer inzicht biedt in bestaande patiënt-gerapporteerde morbiditeitsvragenlijsten en in de validatiestudies. Zijn onderzoek geeft in elk geval een overzicht van bestaande meetinstrumenten. "Maar er is meer onderzoek nodig om te achterhalen of deze vragenlijsten een betrouwbare databron kunnen vormen voor de registratie van morbiditeit."

Het onderzoek werd begeleid door Regan Bergmark, KNO-arts aan de Harvard Medical School en promotor Jan Hazelzet, hoogleraar in het Erasmus MC.

ACKNOWLEDGMENTS / DANKWOORD

"Each of us has cause to think with deep gratitude of those who have lighted the flame within us."
Albert Schweitzer (1875 – 1965)

Ik ben er volledig van bewust dat de succesvolle afronding van mijn PhD-traject is te danken aan vele mensen die, direct of indirect, een belangrijke rol hebben gespeeld voor en tijdens deze fantastische 3.5 jaren. In waarschijnlijk de meest gelezen paragraaf van een proefschrift, verdienen een aantal mensen zwart op wit een speciale "shout-out":

Prof. J.A. Hazelzet, Beste Jan ("prof"), dank voor het vertrouwen dat je in mij had gesteld vanaf het begin van het promotietraject. Toen jij mij aannam als PhD-student had ik nog geen publicaties op mijn naam staan, maar desondanks nam jij een kans op mij en daar ben ik altijd bewust van gebleven. Ik kon altijd spontaan langskomen op jouw kamer met vragen, verzoeken en ideeën. Jij hebt mij veel vrijheid, zelfstandigheid, en een aantal mooie kansen gegeven tijdens het PhD-traject: een leerrijke onderzoeksstage in Boston en de opleiding "Clinical Epidemiology" aan de NIHES. Eén van de belangrijkste adviezen gaf jij mij heel vaak op de vrijdagmiddag: "En nu is het genoeg…lekker genieten van het weekend!". Ik beloof dat ik dat vaker zal doen in de toekomst. Ik heb enorm genoten van de afgelopen 3.5 jaren. Thank you, thank you, thank you!

Prof. N.S. Klazinga, Beste Niek – mijn promotor 'op afstand' – hoewel wij elkaar persoonlijk niet vaak zagen, keek ik altijd uit naar onze 'update' meetings over de status van mijn promotietraject. Met enkele zinnen wist jij altijd frisse en originele perspectieven te bieden over onderzoeksvragen, discussiepunten, en de algehele richting van mijn PhD. Enorm veel dank!

Dr. N. van Leeuwen, Lieve Nikki, de toename in mijn zelfvertrouwen over de afgelopen jaren in het doen van onderzoek, is grotendeels aan jou als copromotor te danken. Jij wist altijd met een grap mij te laten relaxen als iets niet ging volgens mijn planning. Wij hebben samen onvergetelijke en hilarische momenten doorgemaakt: het congres in Portugal waarbij de overige sprekers in het Portugees presenteerden (!), jouw eigen PhD verdediging waarna ik mijn zangdebuut maakte in "Parqiet", mijn onverklaarbare vomitus episode in de gang van een Hilton hotel en een indrukwekkende bike-and-hike trip door Oregon's adembenemend landschap.

Veel dank voor de begeleiding en ik ben erg trots dat ik jouw eerste PhD student ben die zal promoveren! Ik kom gauw langs om Keesje een knuffel te geven.

Dr. H.F. Lingsma, Lieve Hester, mijn onofficiële copromotor. Tijdens mijn sollicitatiegesprek wist ik al dat ik als simpele dokter en onervaren onderzoeker veel zou kunnen leren van jou. Met jouw supersnel inzicht en kritische blik ging ik na onze meetings soms met meer werk de deur uit, maar door jouw humor en spontaniteit keek ik met alle plezier uit naar de volgende afspraak. Ik kon altijd op jou rekenen voor scherpe feedback op de methodologie in mijn manuscripten. Ik hoop dat wij in de toekomst zullen blijven samenwerken aan leuke projecten (liever chirurgisch dan neurologisch). Jou in je element zien op de befaamde Nike campus track in Portland was trouwens onvergetelijk!

Dr. L.B. Koppert, Lieve Linetta, vrij snel nadat ik de baan als PhD-student had gekregen, had ik gehoord dat jij (ongevraagd!) een goed woordje voor mij had gedaan. Dank voor jouw vertrouwen in mij vanaf het begin! Vanaf onze eerste ontmoeting (in de keuken op de afdeling Chirurgische Oncologie in het Daniel den Hoed in 2015), heb ik jou gezien als een voorbeeldfiguur: een chirurg met veel passie, doorzettingsvermogen, nieuwsgierigheid en dat alles altijd met een glimlach. Sindsdien heb ik veel van jou geleerd (o.a. over het mammacarcinoom, de translatie van studieresultaten naar de kliniek, en wat een "flat white" is), hebben wij een aantal mooie artikelen samen gepubliceerd, hebben wij beiden de "Zilveren Kreeft" gewonnen, en ben jij een mentor voor mij geworden! Linet, het is een enorm voorrecht om met jou te mogen samenwerken.

Prof. dr. Sabine Siesling, prof. dr. Hanneke Takkenberg, en dr. Paul Nederkoorn, hartelijk bedankt dat u zitting hebt willen meenemen in de leescommissie. Ik waardeer enorm uw interesse, tijd, en inspanningen voor het beoordelen van mijn proefschrift. De **leden van de commissie** wil ik danken voor het plaatsnemen in de oppositie en het verdiepen in mijn proefschrift.

Looking back on an unforgettable three-month internship at the International Consortium for Health Outcomes Measurement in Cambridge, Massachusetts, USA: **Dr. C. Åkerman, Dear Christina**, thank you very much for extending a warm invitation to briefly join ICHOM in 2018. Ironically, it happened to coincide with the winter season, in which I experienced blizzards and severely freezing temperatures for the first time in my life. Thanks to **Mark, Stephen, Stacie, Sarah, Kemi, Nishwant** and **Liz** for making me feel so welcome; I thoroughly enjoyed all our

meals together, an escape room which ended 'tragically', and of course the movie night at the office.

To all co-authors, both abroad and in the Netherlands, I am incredibly grateful to you for the swift, insightful and valuable input on the manuscripts. I thoroughly enjoyed the collaboration.

Dr. A. Pusic, Dear Andrea, I'm incredibly honored and thankful that you were willing to co-author two articles. I greatly admire the quality of your research, and your all around great personality. I hope we'll actually get to meet in person very soon! **Dr. R. Bergmark, Dear Regan**, thanks for your encouraging supervision and inspiration at ICHOM and onwards, which resulted in a publication I'm very proud of. Your ability to combine clinical care and research is inspiring! **Prof. Dr. D. Dippel, Beste Diederik,** bedankt voor jouw waardevolle inzichten bij 2 manuscripten. Jouw comments hadden altijd een vleugje humor en waren een genot om te lezen. **Prof. Dr. M. Mureau, Beste Marc**, bedankt voor jouw scherpe inzichten en "don't-beat-around-the-bush" attitude!

Pien, bedankt voor jouw inzet tijdens het schrijven van onze uitputtende maar impactvolle systematic review. Veel succes met de opleiding! **Marzyeh**, sincerest apologies from this perfectionist for bothering you with countless questions and feedback during our collaboration on a very insightful paper. Good luck with completing your PhD! **Marloes**, bedankt voor de leuke en vlotte samenwerking aan een publicatie waar ik erg trots op ben. Dat wij meer dan 10 000 vrouwen hebben kunnen bereiken voor de normatieve BREAST-Q studie is iets waar jij even verrast over was als ik!

Bedankt aan de spontane en hulpvaardige secretaresses van de afdeling MGZ! **Marieke**, dank voor de gezellige praatjes en super leuk dat je mijn Broadway tip had opgevolgd. Sugar, butter, flour! **Anja**, bedankt voor de gezellige tijden in de publieke ruimte van Na-24. Jij en ik deelden een appreciatie voor een vaste werkplek. Sorry dat je mijn continu gezang en zittende dansmoves moest tolereren. **Judith**, ik kan niet omschrijven hoe dankbaar ik aan jou ben voor alle ondersteuning gedurende de afgelopen 3.5 jaren. Jij stond altijd, met de liefste glimlach, voor mij en mijn (herhaaldelijke) vragen en verzoeken klaar. Het feit dat mijn proefschrift eerder af is dan verwacht, is ook aan jou te danken!

To all my **colleagues of the department of Public Health**, thanks for the interesting talks, lunches, volleyball games, annual outings and for creating a great

working environment. Lieve **Famke** en **Timor**, jullie hadden mij vanaf dag één welkom laten voelen in Na-2309. Hoewel onze paden iets minder vaak kruisten na de invoering van de flexwerkplekken, wisten wij elkaar altijd te vinden.

Thanks to the **JVO ladies** for their collaboration and enthusiasm in organizing thought-provoking lectures and supporting new PhD students in the department. **Nancy** en **Berivan**, bedankt voor het gezelschap op de "scanty" spookafdeling op de 24e tijdens de COVID-19 crisis.

Thanks to my **CMB colleagues** (**Ana, Benjamin, Charlie, Dana, Daphne, Esmée, Ernest, Evelien, Inge, Iris, Isabel, Jilske, Maaike, Ursula, Vicky**) for the light-hearted but very educational research meetings and hilarious teambuilding events (from glow golfing and laser gaming to building pitiful catapults). **David van Klaveren**, **David** Bedankt voor het delen van jouw kennis over prognostische modellen en voor de potentiele Salmonella blootstelling tijdens het inzetten van voornoemde katapult. **Daan Nieboer**, **Daan,** onze "stats wizard", ik kon altijd op jou rekenen om de bugs in mijn R-scripten te vinden. Super bedankt!

Prof. C. Verhoef en **dr. J.W.A. Burger, Beste Kees** en **Pim**, bedankt voor de kans die jullie mij gaven als geneeskundestudent om kennis te maken met het opzetten van een database en het uitvoeren van klinisch onderzoek. **Kees**, je bent altijd één van mijn grootste supporters geweest. Bedankt voor jouw waardevolle inbreng in dit proefschrift. **Dr. E.V.E. Madsen, Beste Eva**, bedankt voor ook jouw supervisie inclusief gedetailleerde (en snelle!) feedback. Ik ben erg trots dat wij de initiële HIPEC patiënten van het Erasmus MC in kaart hebben kunnen brengen in 2 mooie publicaties!

Erasmus MC, afdeling Heelkunde; Chirurgen, arts-assistenten, verpleegkundig specialisten, verpleegkundigen, secretaresses. Sinds mijn start in 2010 als medisch student in het studententeam op 8-Noord heb ik mij thuis gevoeld op de afdeling. Als ANIOS, kersvers uit de collegebanken, heb ik enorm veel geleerd van jullie en vond ik mijn zelfverzekerdheid als dokter. Ik kon mij geen betere eerste baan hebben voorgesteld!

Hagaziekenhuis, afdeling Heelkunde; Beste collega's, bedankt voor de kennis en gezelligheid vanaf dag één. Wat een team spirit! Ook dank voor jullie bemoedigende woorden tijdens de laatste loodjes van mijn proefschrift.

As writing is an integral part of completing a PhD, a special thank you to my high school English teacher, **Mrs. Rose Simson**. A teacher who nourished my love for the English language: the sonnets of Shakespeare, the metaphysical wit of John Donne, …. "Kindr'd spirit, I hope thee shall see thy owneth lessons being reflected backeth to thee whilst reading this diss'rtation."

Mati and Carina, thank you so much for your talent, passion and time in designing a beautiful and personal cover for my dissertation. It really means a lot!

Mijn paranimfen, **Tirzah** en **Natanael, Tirzie** en **Natty**, wij kennen elkaar inmiddels 19 jaren en jullie hebben mij altijd gesteund in de belangrijke momenten in mijn leven. Ik ben zo blij dat jullie ook fysiek achter mij zullen staan op de "grote dag". **Ping,** bedankt voor jouw zorgzaamheid en luisterend oor. Ik kijk nu al uit naar jouw verdediging in Utrecht. **Natty**, dat je helemaal uit Suriname bent overgevlogen is weer een bewijs van onze hechte vriendschap. Ik waardeer jou enorm, ebbe. Veel succes met jouw opleiding tot kinderarts!

Waar zou ik zijn zonder mijn fantastische vrienden? Sommigen van jullie ken ik al praktisch mijn hele leven en zijn inmiddels broers en zussen geworden! Bedankt aan alle lieve vrienden in Suriname (**Natanael, Xavier T., Zaïra, Serena, Devika**), Nederland **(Chandèl, Xaviera, Tirzah, Marit, June, Stephan, Rick, Jaco, Nathalie, Ananta, Priscilla, Ceranza, Jeroen, Jayshri**) en Berlijn (**Xavier R.**) voor hun voortdurende support, geduld, vertrouwen, humor, "shade", gezelligheid, en afleiding! Ik kijk uit naar nog meer borrels, dinners, game nights, en vakanties! Jullie accepteren mij zoals ik ben, ondanks mijn vele tekortkomingen. I'm a lucky guy.

Maxie, dat onze paden zijn gekruist aan het einde van mijn promotie was zeker niet bevorderlijk voor een snelle afronding maar ik had het absoluut niet anders gewild. Dank voor jouw humor, spontaniteit, steun, geduld en liefde. Love you, buster!

Mijn dank is groot aan alle **Oemrawsingh, Gangaram Panday, Loor, Beeker, Issa** en **Brahim** familieleden voor hun oprechte interesse en bemoedigende woorden sinds mijn verhuizing naar Nederland in 2008. **Vishal, Renuka en Sujata**, wat bof ik met "cousies" als jullie! Voor eeuwig zal ik jullie dankbaar zijn voor de initiële opvang in Delft toen ik naar het koude kikkerland verhuisde, voor het uitleggen wat het verschil was tussen een stop- en sneltrein, en voor de kooklessen. Mijn zelfstandigheid is het product van jullie ongelofelijke steun. **Marlene phoewa, oom Rama, Sietske phoewa, Perry dada, tante Diana, oom Dhier, tante Agnes**,

hoewel wij meer dan 7500km van elkaar wonen, heb ik jullie liefde en support altijd zeer sterk gevoeld. **Tante Maggie, oom Robert, tante Inez, oom Jerry,** veel dank voor de steun vanuit Almere!

Aan mijn lieve zusjes, **Sarahna, Arishta** en **Kayleigh**; wat ben ik ontzettend trots op jullie! Hardwerkende, integere, zelfstandige en mooie dames, die elk hun eigen pad bewandelen. Ik ben erg blij dat wij, na véle jaren, eindelijk weer in één land wonen! Dat geeft mij namelijk steeds nieuwe kansen om de broer te zijn die jullie verdienen. **Vashu**, eindelijk een broertje erbij om te plagen! Ik wacht nog op een goede comeback van jou...

Allerliefste **papa en mama**, dit proefschrift draag ik op aan jullie. Zie het als een klein bedankje van jullie eerstgeborene voor de onvoorwaardelijke liefde en ondersteuning die jullie hem blijven geven. Pa, wij verschillen in vele opzichten van elkaar, maar ik ben blij dat ik jouw koppigheid en zelfsturend vermogen heb geërfd. Je bent de moedigste man die ik ken. Mother darling, ik kijk ernaar uit om jou nog meer te verwennen met stedentripjes, opera's en etentjes. Ik ben zo blij dat jullie overvliegen om dit feestelijk moment met mij te delen. Hoewel ik het niet vaak genoeg uit, ben ik dagelijks dankbaar voor en bewust van alles dat jullie mij hebben gegeven: een veilig en warm thuis, een luisterend oor en, het allerbelangrijkste, de vrijheid om mijn eigen weg te vinden door het leven. Ik houd van jullie.

A.
Rotterdam, Juli 2021

PHD PORTFOLIO

ERASMUS UNIVERSITEIT ROTTERDAM

Name PhD student:	Arvind Oemrawsingh
Erasmus MC Department:	Public Health – Section Medical Decision Making
PhD period:	May 2017 – October 2020
Promotors:	Prof. Dr. J.A. Hazelzet
	Prof. Dr. N.S. Klazinga
Copromotor:	Dr. N. van Leeuwen

PhD Training	**Year**	**Workload (ECTS)**
MHS in Clinical Epidemiology, Netherlands Institute for Health Sciences (NIHES), Rotterdam, the Netherlands	2017-2019	70.9
Core Curriculum		
Study Design (CC01)	2017	4.3
Clinical Epidemiology (CE02)	2017	3.7
Clinical Translation to Epidemiology (CE01)	2017	2.0
Biostatistical Methods I: Basic Principles (CC02)	2018	5.7
Biostatistical Methods II: Classical Regression Models (EP03)	2018	4.3
Principles in Causal Inference (EP01)	2018	1.4
Introduction to Medical Writing (SC02)	2018	2.0
Research Paper (M-RES)	2018	34.0
Advanced Elective Courses		
Missing Values in Clinical Research (EP16)	2018	1.4
Advanced Topics in Decision-Making in Medicine (EWP02)	2019	2.4
Quality of Life Measurement (HS11)	2019	0.9
Planning and Evaluation of Screening (HS05)	2019	1.4
Cancer Epidemiology (EP13)	2019	1.4
From Problem to Solution in Public Health (HS18)	2019	1.1
Erasmus Summer Programme		
Principles of Research in Medicine and Epidemiology (ESP01)	2017	0.7
Clinical Trials (ESP14)	2017	0.7
Methods of Public Health Research (ESP11)	2017	0.7
Fundamentals of Medical Decision Making (ESP70)	2017	0.7
The Practice of Epidemiologic Analysis (ESP65)	2018	0.7
Health Economics (ESP25)	2018	0.7
Methods of Health Services Research (ESP42)	2018	0.7

General Courses		
CPO Course - Patient Oriented Research: design, conduct and analysis	2017	0.3
Systematic Literature Retrieval in PubMed, Medical Library, Erasmus MC, Rotterdam, the Netherlands	2017	0.3
Endnote, Medical Library, Erasmus MC, Rotterdam, the Netherlands	2017	0.3
Scientific Integrity, Erasmus MC, Rotterdam, the Netherlands	2019	0.3
Causal Inference and Causal Diagrams in Medical Decision Making Using Big Real World Observations, Society for Medical Decision Making (SMDM)	2019	0.3
Poster Presentations		
Netherlands Epidemiological (WEON) Conference, Bilthoven, the Netherlands	2018	1.0
International Consortium of Health Outcomes Measurement (ICHOM) Conference, Rotterdam, the Netherlands	2019	1.0
41st Annual North American Meeting, Society for Medical Decision Making (SMDM), Portland, Oregon, USA	2019	1.0
European Breast Cancer Conference (EBCC 12) – Virtual	2020	1.0
Oral Presentations (Conferences, Symposia and Seminars)		
ACE Health Care Quality, Erasmus MC, Rotterdam, the Netherlands	2017	1.0
17th Biennial European Conference with Special Focus on Personalized and Value-Based Healthcare, Society for Medical Decision Making (SMDM), Leiden, the Netherlands	2018	1.0
Value-Based Healthcare (VBHC) Research Meeting, Department of Public Health, Erasmus MC, Rotterdam, the Netherlands	2018	1.0
Research Seminar, Department of Public Health, Erasmus MC, Rotterdam, the Netherlands	2019	1.0
Research in Value-Based Healthcare, Nova School of Business & Economics, Lisbon, Portugal	2019	1.0
Smarter Choices for Better Health, Erasmus University Rotterdam, the Netherlands	2019	1.0
Wetenschapsdag Heelkunde, Erasmus MC, Rotterdam, the Netherlands	2020	1.0
Supervising/ Teaching Activities		
Vaardigheidsonderwijs (VO) “Hoe houden we de zorg betaalbaar?” to 3rd-year medical students	2019	1.0
Lecture “Measuring Quality of Life: Using Patient-Reported Outcome Measures (PROMs) in Clinical Practice” to 3rd-year BMG (Beleid & Management Gezondheidszorg) students	2019	1.0
Supervising the Community Project “Verschillen in kwaliteit van leven tussen gezonde vrouwen en borstkanker patiënten” of 3rd-year medical students	2020	1.5
Master Consultancy project: Assisting medical students with statistical analyses during their MSc. research project	2020	1.0
Research Internship		
Research Assistant at International Consortium for Health Outcomes Measurement (ICHOM), Cambridge, Massachusetts, USA	2018 (Jan-April)	21.0

Organizational Activities		
Organizing Committee for annual MGZ-day, Department of Public Health, Erasmus MC, Rotterdam, the Netherlands	2018	
Organizing Committee for annual "Medical Decision Making" section outing, Department of Public Health, Erasmus MC, Rotterdam, the Netherlands	2019	
Memberships		
Member of Junior Representatives Committee (JVO), Department of Public Health, Erasmus MC, Rotterdam, the Netherlands	2018-2020	
Member of the European University Hospital Alliance (EUHA) – Breast Cancer Working Group	2019	
Awards		
"De Zilveren Kreeft" Bokaal, Wetenschapsdag Heelkunde, Erasmus MC, Rotterdam, the Netherlands Best Abstract & Oral Presentation: "Longitudinal Patient-Reported Outcomes Within Breast Cancer Surgical Treatment Strategies"	2020	

ABOUT THE AUTHOR

Arvind Oemrawsingh was born on March 7th, 1990 in Plantation, Florida, USA. He grew up in Paramaribo, Suriname. In 2008, he received his high school degree from the Arthur Alex Hoogendoorn Atheneum.

Afterwards, he moved to the Netherlands to study Medicine at the Erasmus University Rotterdam. During this period, he was involved in many student committees which organized lectures and workshops for students. He also worked for over three years at the Department of Gastro-Intestinal Surgery as a medical student, where he assisted the nursing staff in specialized surgical care. A couple of medical internships were also spent abroad during this time: Emergency Medicine in Egypt (2010) and Surgery in Suriname (2013).

During his surgical internship at Maasstad Hospital, he discovered a passion for surgery, and decided to do a research internship at the Department of Surgical Oncology at the Erasmus MC Cancer Centre, focusing on the short-term postoperative course after cytoreductive surgery and hyperthermic intraperitoneal chemotherapy (CRS + HIPEC) in patients with peritoneal carcinomatosis (supervisors: prof. dr. C. Verhoef, dr. J.W.A. Burger and dr. E.V.E. Madsen). After receiving his medical degree, he started 2016 off as a surgical resident (ANIOS) at the Erasmus MC.

In April 2017, he got the opportunity to begin as a PhD candidate with a focus on patient-reported outcome measures (promotors: prof. dr. J.A. Hazelzet & prof. dr. N.S. Klazinga), as part of the Value-Based Healthcare program of the Dutch Federation of University Medical Centers (NFU). Simultaneously, he started a second MSc. program in Clinical Epidemiology at the Netherlands Institute for Health Sciences (NIHES), which he completed in August 2019. As a first-year PhD-student, he was invited to spend three months as a research assistant at the International Consortium for Health Outcomes Measurement (ICHOM) in Boston, Massachusetts, USA. During his PhD training, Arvind was also a member of the

Junior Representatives' Committee (JVO) and supervised medical students with their own research projects.

Arvind, a fierce lover of the performing arts, enjoys playing the piano and singing as a tenor in the Luxor theatre choir in Rotterdam. In his spare time, he has also been volunteering as a COC (Cultureel Ontspanningscentrum) lecturer at elementary schools in Amsterdam, talking to students about LGBT-related topics, encouraging acceptance and battling school bullying.

Arvind is currently working as a surgical resident (ANIOS) in the Hagaziekenhuis.

PROPOSITIONS BELONGING TO THE THESIS

Patient-Reported Outcome Measures in Clinical Care: Methodological Aspects and Practical Applications

1) Patient-reported morbidity instruments can be a reliable data source depending on the included items. *This thesis*
2) Case-mix adjustment models for patient-reported outcome measures are different from those for clinical outcomes when applied to ischemic stroke care. *This thesis*
3) The extent to which patient-reported outcome measures are influenced by traditional baseline patient characteristics can differ between domains of quality of life. *This thesis*
4) The trend of patient-reported outcome measurement scores during breast cancer treatment and follow-up may help both patients and clinicians to better understand treatment impact. *This thesis*
5) For a successful implementation of patient-reported outcome measures in clinical care, healthcare institutions should solicit knowledge on IT-infrastructure from experienced hospitals rather than reinvent the technological wheel themselves. *This thesis*
6) Providers, payers, patient-advocacy groups, and regulators can come together to create a process to agree on a minimum sufficient set of outcomes for each important medical condition — including rigorous definitions, risk-adjustment factors, and methods. *Michael Porter et al., New England Journal of Medicine, 2016*
7) As developing technology continues to increase our capacity to measure, store, and analyze larger amounts of data, it is important to consider parsimony of data collection. *Lee Squitieri et al., Value in Health, 2017*
8) Outcomes, by and large, remain the ultimate validators of the effectiveness and quality of medical care. *Avedis Donabedian, The Milbank Quarterly, 2005*
9) Extensive research in disaster mental health has established that emotional distress is ubiquitous in affected populations — a finding certain to be echoed in populations affected by the COVID-19 pandemic. *Betty Pfefferbaum et al., New England Journal of Medicine, 2020*
10) Humanity's greatest advances are not in its discoveries – but in how those discoveries are applied to reduce inequity. *Bill Gates, Commencement Address at Harvard University in Cambridge, MA, June 7, 2007*
11) Anything you do, let it come from you. Then it will be new. *Stephen Sondheim, Sunday in the Park with George, 1984*

Arvind Oemrawsingh

1 oktober 2021