

DISSERTATION

Vergleichbarkeit von Patient-Reported Outcome Measures:
Zwei Ansätze zur Standardisierung

Comparability of patient-reported outcome measures:
Two approaches for standardization

zur Erlangung des akademischen Grades
Doctor rerum medicinalium (Dr. rer. medic.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Claudia Hartmann

Erstbetreuung: Prof. Dr. med. Matthias Rose

Datum der Promotion: 28.02.2025

Inhaltsverzeichnis

Abbildungs- und Tabellenverzeichnis	i
Abkürzungsverzeichnis.....	ii
Zusammenfassung	1
1 Einleitung.....	4
2 Methodik.....	7
2.1 Publikation 1: Das ICHOM Standard Set for Pregnancy and Childbirth – Translation and Linguistic Adaption for Germany.....	8
2.1 Publikation 2: PROMIS-29 and EORTC QLQ-C30 – an empirical investigation towards a common conception of health	9
3 Ergebnisse	12
Publikation 1: Das ICHOM Standard Set for Pregnancy and Childbirth – Translation and Linguistic Adaption for Germany	12
Publikation 2: PROMIS-29 and EORTC QLQ-C30 – an empirical investigation towards a common conception of health.....	13
4 Diskussion und Implikationen	16
Literaturverzeichnis	19
Eidesstattliche Versicherung	22
Anteilerklärung an den erfolgten Publikationen.....	23
Auszug aus der Journal Summary List	24
Druckexemplar(e) der Publikation(en).....	25
Lebenslauf.....	43
Komplette Publikationsliste.....	44
Danksagung	45

Abbildungs- und Tabellenverzeichnis

Abbildung 1: Arbeitsschritte FACIT-Methode.....	9
Tabelle 1: Skalen des EORTC QLQ-C30.....	10
Abbildung 2: CFA-Modell mit 6 Domänen und Ladungen	15

Abkürzungsverzeichnis

BSES-SF	Breastfeeding Self-Efficacy Scale - Short Form
BSS-R	Birth Satisfaction Scale – Revised
CFA	konfirmatorische Faktorenanalyse
CFI	Comparative Fit Index
COMET	Core Outcome Measures in Effectiveness Trials Initiative
EFA	explorative Faktorenanalyse
EORTC	European Organisation for Research and Treatment of Cancer
FACIT	Functional Assessment of Chronic Illness Therapy
FDA	Food and Drug Administration
ICHOM	International Consortium for Health Outcomes Measurement
IRT	Item-Response-Theorie
MIBS	Mother-to-Infant Bonding Scale
PROMIS	Patient-Reported Outcomes Measurement Information System
PROMs	Patient-Reported Outcomes Measures
PROs	Patient-Reported Outcomes
RMSEA	Root Mean Square Error of Approximation
SONG	Standardised Outcomes in Nephrology
SRMR	Standardized Root Mean Square Residual

Zusammenfassung

Einleitung: Patient-Reported Outcomes (PROs) werden zunehmend in der Routineversorgung eingesetzt. Sie werden als Qualitätsindikatoren in ein erweitertes Qualitätsverständnis integriert und dienen als Instrument einer patientenzentrierten Medizin. In den letzten Jahrzehnten wurde eine Vielzahl von PRO Measures (PROMs) entwickelt, wodurch die Frage der Vergleichbarkeit von PRO-Messungen und somit der Standardisierung an Relevanz gewonnen hat. Aktuell gibt es hierzu zwei führende Ansätze: Die Standardisierung durch die Nutzung gleicher Instrumente und die Standardisierung mittels Einigung auf eine gemeinsame Skala/Metrik. Ziel der Arbeit ist es, zwei konkrete Beiträge zur Etablierung von Standards zu leisten, um die derzeitige Vergleichbarkeit von PRO-Ergebnisse zu verbessern.

Methoden: Im ersten Teil der Arbeit (1) wird das Standardset für Schwangerschaft und Geburt des International Consortium for Health Outcomes Measurement (ICHOM) mit den Instrumenten: Birth Satisfaction Scale – Revised (BSS-R) Breastfeeding Self-Efficacy Scale - Short Form (BSES-SF), Mother-to-Infant Bonding Scale (MIBS) und 16 Einzelitems nach der Functional Assessment of Chronic Illness Therapy Methode (FACIT) ins Deutsche übersetzt und validiert. Im zweiten Teil (2) wird eine Grundlage für den Einsatz einer einheitlichen Metrik geleistet, in dem mittels Inhaltsanalyse, explorativer und konfirmatorischer Faktoranalyse das zugrundeliegende Konstrukt von Gesundheit der Instrumente PROMIS-29 und QLQ-C30 untersucht und verglichen wird.

Ergebnisse: (1) Alle Items und Instrumente konnten mit einem hohen Maß an Verständlichkeit übersetzt werden. Es zeigte sich, dass medizinische Begriffe im Sprachgebrauch nicht immer eindeutig interpretiert werden, weshalb auf eine klare und fachlich korrekte Formulierung geachtet wurde. Die Antwortstruktur sowie der Itemstamm des BSES-SF wurden angepasst, um eine einheitliche und verständliche Übersetzungslogik zu erhalten. (2) Die Korrelationsmatrix zwischen PROMIS-29 und QLQ-C30 zeigt Werte zwischen 0,59 (Soziale Rolle) und 0,78 (Körperliche Funktion). Das entwickelte CFA-Modell mit 6 Domänen weist eine hinreichende Passung der Daten auf (RMSEA = 0,074, CFI = 0,968 und SRMR = 0,074). Die durchgeführten Analysen unterstützen die Annahme einer konzeptionellen Übereinstimmung beider Instrumente.

Schlussfolgerung: Standardisierung in der PRO-Messung ist zwingend erforderlich. Derzeit werden im Wesentlichen zwei Ansätze verfolgt. Mit der Arbeit steht der Geburtsmedizin in Deutschland nun ein international konzertiertes Set mit wichtigen PRO-Instrumenten für international vergleichende Analysen zur Verfügung. Darüber hinaus legt die Arbeit die Grundlage für die Entwicklung eines Transformationsalgorithmus oder der Kalibrierung auf eine einheitliche Metrik für die Instrumente PROMIS-29 und QLQ-C30, indem sie zeigt, dass die zugrundeliegende Konzeptualisierung von Gesundheit hinreichend ähnlich ist.

Abstract

Introduction: Patient-reported outcomes (PROs) are increasingly being used in routine care. They are integrated as quality indicators into an expanded understanding of quality and serve as an instrument for patient-centered medicine. In the last few decades, a large number of PRO measures (PROMs) have been developed, which means that the question of comparability of PRO measurements and thus standardization has become more relevant. There are currently two leading approaches to this: standardization by using the same instruments and standardization by agreeing on a common scale/metric. The aim of the work is to have two concrete contributions to the establishment of standards for the improvement of the comparability of PRO results.

Methods: In the first part of the work (1), the standard set for obstetrics of the International Consortium for Health Outcomes Measurement (ICHOM) is translated into German and validated for the following instruments: Birth Satisfaction Scale - Revised (BSS-R) Breastfeeding Self-Efficacy Scale - Short Form (BSES-SF), Mother-to-Infant Bonding Scale (MIBS) and 16 individual items using the Functional Assessment of Chronic Illness Therapy (FACIT) method. In the second part (2), an important foundation is laid for the use of a uniform metric by investigating and comparing the underlying construct "general health" of the instruments PROMIS-29 and QLQ-C30 using content analysis, exploratory and confirmatory factor analysis.

Results: (1) All items and instruments could be translated with a high degree of comprehensibility. It became apparent that medical terms are not always interpreted

unambiguously in everyday language, which is why attention was paid to clear and technically correct wording. The answer structure and item pool of the BSES-SF were adapted to maintain a uniform and comprehensible translation logic. (2) The correlation matrix between PROMIS-29 and QLQ-C30 shows values between 0.59 (social role) and 0.78 (physical function). A CFA model with 6 domains was developed, which shows a sufficient fit of the data (RMSEA = 0.074, CFI = 0.968 and SRMR = 0.074). The analyses performed support the assumption of a conceptual agreement between the two instruments.

Conclusion: Standardization in PRO measurement is essential. Currently, two main approaches are being pursued. With this work, 3 important international PRO measures are now available for obstetrics in Germany and enable comparative international analyses. In addition, by demonstrating that the underlying conceptualization of health is sufficiently similar, the work lays the groundwork for developing comparative algorithms or calibration to a common metric for the PROMIS-29 and QLQ-C30 instruments.

1 Einleitung

1.1 Thematische Einordnung

Patient-Reported Outcomes (PROs) erhalten zunehmend einen zentralen Platz in der Gesundheitsversorgung und bei der Gestaltung des Gesundheitssystems. Sie erweitern den Qualitätsbegriff der Medizin um die Patientenperspektive und ermöglichen so eine ganzheitliche Bewertung medizinischer Behandlungen.[1] Dies stellt einen entscheidenden Baustein für den Paradigmenwechsel hin zu einer wertebasierten Gesundheitsversorgung (Value-Based Healthcare) dar. Dieser Wandel zielt auf eine bessere Verteilung der Kosten, effizientere Ressourcennutzung und qualitativ hochwertige medizinische Leistungen ab. Im Zentrum steht die Definition des Wertes einer Behandlung als Verhältnis zwischen den erzielten Ergebnissen (Outcomes) und den dafür aufgetragenen Kosten.[2] Daneben spielen PROs eine immer größere Rolle in der Routineversorgung. Die ätiologisch ausgerichtete Medizin, die auf die Heilung von Krankheiten und die Erhöhung der Lebenserwartung abzielt, befindet sich an einem Wendepunkt. Die Zunahme chronischer Erkrankungen und die Stagnation der Lebenserwartung erfordern eine Verschiebung der Versorgung hin zu einem Management des Lebens mit Krankheit. In diesem Kontext helfen PROs Krankheitssymptome und Nebenwirkungen im Behandlungsprozess individuell für Patient:innen zu bewerten und die Therapie bei Bedarf anzupassen.[3]

Die wachsende Bedeutung von PROs wird auch auf regulatorischer und politischer Ebene deutlich. Die US-amerikanische Food and Drug Administration (FDA) hatte 2009 die Methodik für die Berücksichtigung von PROs in der Nutzenbewertung neuer Medikamente und Behandlungsmethoden vorgeschlagen.[4] Im selben Jahr begann in England die systematische Erhebung von PROs für den Bereich Knie- und Hüftendoprothetik.[5] Teil der nationalen Gesundheitspolitik sind PROs mittlerweile u.a. in Schweden, den Niederlanden und Dänemark.[6-8] In Deutschland finden PROs derzeit vor allem im Rahmen verschiedener Pilot- und Forschungsprojekte einzelner Kliniken oder Initiativen Anwendung. Die Erhebung von PROs gewinnt vor allem an Bedeutung, da sie vermehrt im Kontext von Qualitätsverträgen gemäß §110 des Sozialgesetzbuches V (SGB V) eingeführt werden. Von einer breiten, standardisierten Nutzung der PROs in Deutschland kann allgemein jedoch nicht gesprochen werden.[9]

PRO Measures werden bereits seit den 1970er entwickelt und erforscht.[10] Die weltweit wachsende Bedeutung hat dazu geführt, dass immer mehr und spezifischere PRO Measures (PROMs) entwickelt werden. So finden sich auf der Seite von Mapi Trust, einer Onlinedatenbank für PROMs, alleine 1.442 Instrumente zum Begriff allgemeine Gesundheit (general health) und 289 zum Thema Depression (Stand: 01.05.2024).[11] Die hohe Diversität zeigt sich auch in Metastudien für verschiedene Krankheitsbilder.[12-14] Aufgrund der unterschiedlichen Skalen und fehlender Vergleichsmethoden ist es meist nicht möglich, vergleichende Analysen oder Metaanalysen durchzuführen. Standardisierung ist daher ein dringendes Thema, möchte man sich zukünftig zu PRO-Daten international austauschen und vergleichen. Aktuell gibt es hierzu zwei führende Ansätze: Die Standardisierung durch die Nutzung gleicher Instrumente und die Standardisierung mittels Einigung auf eine gemeinsame Ergebnis-Skala/-Metrik.

Den Ansatz ein standardisiertes Instrumentenset zu definieren, verfolgen eine Vielzahl an Organisationen. So gibt es Initiativen für einzelne Fachbereiche z.B. Standardised Outcomes in Nephrology (SONG) oder die European Organisation for Research and Treatment of Cancer (EORTC) und für die wissenschaftliche PRO-Erhebung z.B. Core Outcome Measures in Effectiveness Trials Initiative (COMET). Eine Organisation die den Fokus auf die Entwicklung und Verbreitung von Standardsets in der klinischen Routineerhebung legt, ist das International Consortium for Health Outcome Measurement (ICHOM), welches 2012 als gemeinnützige Organisation gegründet wurde. Die Initiative entwickelt Standardsets für verschiedene Krankheitsbilder und Populationen, die neben bereits validierten PROMs auch klinische Ergebnisdaten (z.B. Laborergebnisse, Tumormarker etc.) und soziodemografische Case-Mix-Faktoren beinhalten. Mittlerweile wurden 46 Standardsets entwickelt (Stand: 01.05.2024).[15; 16]

Eine Vergleichbarkeit der PRO-Ergebnisse kann aber auch über die Festlegung einer gemeinsamen Metrik erfolgen. Dieser Ansatz ermöglicht weiterhin die Nutzung individuell präferierter Fragebögen, fordert jedoch die Darstellung der Ergebnisse auf einer einheitlichen und nicht auf einer instrumentenspezifischen Skala. Eine passende Metrik kann hierfür die T-Metrik mit einem Skalenwert (T-Score) von 0-100 (theoretisch $\pm \sim$) sein, welche auf Basis der Item-Response-Theorie (IRT) entwickelt wurden. Auf dieser Metrik stellt 50 den Populationsmittelwert da. Mit einer Standardabweichung von ± 10 wird der

Normbereich definiert. Im Rahmen dieses Standardisierungsansatzes werden Umrechnungsalgorithmen entwickelt oder Instrumente auf die neue Metrik kalibriert. Das Patient-Reported Outcomes Measurement Information System (PROMIS®) ist die führende Initiative für PROMs, die auf IRT-Modellen basieren.[17; 18]

1.2 Zielstellung

Ziel der Arbeit ist es, zwei konkrete Beiträge zur Etablierung von Standards gemäß den oben erläuterten Ansätzen zu leisten, um die Vergleichbarkeit von PRO-Ergebnissen zu verbessern.

- 1) Übersetzung der bisher nicht auf Deutsch verfügbaren PROMs und Einzelitems des ICHOM Standard Set für Schwangerschaft und Geburt auf Deutsch inkl. sprachlicher Validierung.
- 2) Vergleich der Instrumente PROMIS-29 und EORTC QLQ-C30 hinsichtlich ihrer zugrundeliegenden Konzeptionalisierung von Gesundheit, um zukünftig einen direkten Vergleich beider Outcomes auf einer Metrik zu ermöglichen.

2 Methodik

Als PROs werden alle Informationen über den Gesundheitszustand von Patient:innen verstanden, welche direkt von ihnen berichtet werden und nicht von einer dritten Person gedeutet werden müssen.[19] In Anlehnung an die WHO Definition von Gesundheit, beschreiben Informationen über den allgemeinen Gesundheitszustand körperliche, psychische und soziale Faktoren.[20] Je nach Krankheitsbild wird den Faktoren unterschiedlich starke Bedeutung zugemessen und ggf. kommen auch noch therapie- oder krankheitsspezifische dazu. Die einzelnen Faktoren gliedern sich wiederum in verschiedene latente Konstrukte auf, z.B. Depressivität, Angst, körperliche Aktivität, Schmerz, soziale Teilhabe und viele mehr.[21]

Zur Erfassung von PROs werden Instrumente oder Fragebögen – sogenannten Patient-Reported Outcomes Measures (PROMs) genutzt. Diese können in zwei Hauptkategorien unterteilt werden: generische und krankheitsspezifische. Generische PROMs sind allgemeine Fragebögen, die auf eine breite Palette von Gesundheitszuständen und Patientenpopulationen zutreffen. Sie ermöglichen es, die Lebensqualität und das Wohlbefinden unabhängig von einer spezifischen Krankheit oder Behandlung zu bewerten. Ein Beispiel hierfür ist der PROMIS-29. Im Gegensatz dazu sind krankheitsspezifische PROMs speziell darauf ausgerichtet, den Einfluss einer bestimmten Krankheit oder Behandlung auf die Gesundheit und Lebensqualität der Patient:innen zu messen. Diese Instrumente sind enger auf die Symptome, Funktionen und Bedürfnisse im Zusammenhang mit einer bestimmten Erkrankung abgestimmt und liefern detailliertere Einblicke in den Krankheitsverlauf und die Behandlungseffektivität. Beispiele hierfür sind der Expanded Prostate Cancer Index Composite mit 26 Items (EPIC-26) für Patienten mit Prostata Karzinom oder die Breastfeeding Self-Efficacy Scale (BSES) in der Geburtsmedizin.[21]

2.1 Publikation 1: Das ICHOM Standard Set for Pregnancy and Childbirth – Translation and Linguistic Adaption for Germany

Das ICHOM Standard Set für „Schwangerschaft und Geburt“ wurde von einer internationalen Gruppe bestehend aus Hebammen, Gynäkolog:innen und Patientenvertreterinnen entwickelt. In dem Set werden sowohl medizinische, klinisch berichtete als auch patientenberichtete Daten erhoben. Folgende Themenbereiche werden durch die PROs abgedeckt:

- Allgemeines: Demografie und Anamnese,
- allgemeine Gesundheit und Lebensqualität,
- Geburtserleben und Zufriedenheit mit der Versorgung,
- Stillen,
- Rollenwechsel,
- psychische Gesundheit,
- Inkontinenz sowie,
- Schmerzen beim Geschlechtsverkehr.[22]

Im Set werden dafür 8 Instrumente und 19 Einzelfragen erhoben, wobei 5 Instrumente und 3 Einzelfragen bereits validiert auf Deutsch verfügbar sind. Die noch nicht auf Deutsch übersetzten Instrumente sind die: Birth Satisfaction Scale – Revised (BSS-R), Breastfeeding Self-Efficacy Scale - Short Form (BSES-SF) und die Mother-to-Infant Bonding Scale (MIBS) sowie 16 Einzelfragen aus den Bereichen Demografie und Anamnese (8), Zufriedenheit mit der Versorgung (5), Erfolg beim Stillen (3), Inkontinenz (1) und Schmerzen beim Geschlechtsverkehr (1).[23]

Die Übersetzung und sprachliche Validierung (Cognitive Debriefing) der Instrumente und Items erfolgte nach der Functional Assessment of Chronic Illness Therapy (FACIT) Methode. Die FACIT Methode besteht aus 6 Schritten die in Abbildung 1 dargestellt ist. Im ersten Schritt werden die Fragen aus dem englischen von zwei Personen unabhängig voneinander ins Deutsche übersetzt und anschließend in einer Version konsolidiert. Die deutsche Übersetzung wird dann von zwei Muttersprachler:innen zurück ins Englische übersetzt. Im nächsten Schritt werden die Originalversion und die Rückübersetzung miteinander verglichen und etwaige Abweichungen fachlich geprüft und bei Bedarf in der deutschen Formulierung geändert. Abschließend wird ein Cognitive Debriefing mit der

Zielgruppe durchgeführt.[24] Das Cognitive Debriefing beschreibt dabei ein qualitatives Verfahren zur sprachlichen Validierung der Übersetzung mit dem Ziel Verständnisprobleme und Unklarheiten in der Übersetzung zu identifizieren.[25] Schwangere und Wöchnerinnen wurden in diesem Schritt gebeten die übersetzten Items laut vorzulesen, den Inhalt der Items mit Ihren eigenen Worten wiederzugeben und ihre Gedanken dahingehend zu teilen, wie sie die Items verstehen. Insgesamt wurde in drei Gruppen mit jeweils 5 Frauen getestet. Anschließend wurden die Hinweise und Anmerkungen aus dem Cognitive Debriefing im Team diskutiert und in die Übersetzung eingearbeitet.[26]



Abbildung 1: Arbeitsschritte FACIT-Methode [23]

2.1 Publikation 2: PROMIS-29 and EORTC QLQ-C30 – an empirical investigation towards a common conception of health

Instrumente

Das PROMIS-29 Profil ist ein krankheitsunabhängiges Instrument innerhalb von PROMIS bestehend aus 29 Items. Dieser Kurzfragebogen (Short Form) deckt sieben zentrale Domänen der gesundheitsbezogenen Lebensqualität mit jeweils 4 Items ab und enthält ein Einzelitem zur Schmerzintensität. Die sieben Domains lauten:

- körperliche Funktionsfähigkeit,
- Schlafbeeinträchtigung,
- Beeinträchtigung durch Schmerzen
- Erschöpfung,
- Emotionale Belastung - Angst,
- Emotionale Belastung – Depression,
- Teilnahme an sozialen Rollen und Aktivitäten.[27]

Der QLQ-C30 ist ein zentrales Instrument innerhalb der Sammlung von EORTC, das sich speziell auf die allgemeine gesundheitsbezogene Lebensqualität von Krebspatient:innen konzentriert. Tabelle 1 zeigt, wie sich die 30 Items auf die fünf funktionalen Skalen, acht Symptomskalen und drei Einzelskalen verteilen.[28; 29]

Tabelle 1: Skalen des EORTC QLQ-C30

Anmerkung: eigene Darstellung

Funktionsskalen	Symptomskalen	Weitere Skalen
<ul style="list-style-type: none"> • soziale Funktion • emotionale Funktion • kognitive Funktion • Rollenfunktion • körperliche Funktion 	<ul style="list-style-type: none"> • Schmerzen • Erschöpfung • Atemnot • Schlaflosigkeit • Appetitverlust • Verstopfung • Durchfall • Übelkeit & Erbrechen 	<ul style="list-style-type: none"> • finanziellen Status • globale Lebensqualität • globale Gesundheitsskala

Auswertungen

Im ersten Schritt wurde ein inhaltlicher Vergleich der einzelnen Items beider Instrumente erstellt. So wurden die einzelnen Items hinsichtlich Ihrer Formulierung und Anzahl der Items zu einem Faktor gegenübergestellt.

Für die statistische Analyse wurden Daten aus der Routineversorgung im Brustzentrum der Charité Universitätsmedizin Berlin (Charité) verwendet, wo der PROMIS-29 und der QLQ-C30 erhoben werden. Bei der Erhebung konnten keine Fragen übersprungen werden, weshalb die Erhebung vollständig ist und keine fehlenden Daten enthält. Für die Analyse wurden insgesamt 1.478 Erstbefragung aller Patientinnen im Zeitraum November 2016 bis März 2021 verwendet.

Für jede Subskala wurden die Scores des QLQ-C30 und PROMIS-29 in T-Werte umgewandelt und anschließend Mittelwerte und Standardabweichungen berechnet. Diese Daten wurden mit allgemeinen Bevölkerungsdaten aus früheren Veröffentlichungen verglichen.[30; 31] Die Korrelationen zwischen allen Bereichen wurden mithilfe des Pearson-Korrelationskoeffizienten bewertet, wobei Korrelationen

über 0,6 als hoch und über 0,4 als moderat betrachtet wurden [32]. Darüber hinaus wurde sowohl eine explorative Faktorenanalyse (EFA) als auch eine konfirmatorische Faktorenanalyse (CFA) durchgeführt, um die dimensionale Struktur beider Instrumente zu bewerten. In der Korrelationsanalyse, EFA und CFA wurden die Symptomskalen des QLQ-C30, die globalen Skalen sowie die Items zu kognitiven Funktion a priori ausgeschlossen. Darüber hinaus wurde das PROMIS-29 Einzelitem zur Schmerzintensität ausgeschlossen, da es sich um eine numerische Analogskala handelt.

Für die CFA und EFA wurde der kombinierte Datensatz zufällig in zwei Hälften geteilt, so dass zwei gleich große Stichproben von $N=739$ entstanden. Mehrere Faktorenlösungen wurden explorativ anhand von Faktorladungen untersucht und verglichen.

Für die CFA wurde ein Modell entwickelt, das auf der Analyse deskriptiver Maße und den Ergebnissen der EFA basiert. Die Anpassung des Modells wurde anhand des Comparative Fit Index (CFI), des Root Mean Square Error of Approximation (RMSEA) und des Standardized Root Mean Square Residual (SRMR) bewertet, wobei die Kriterien für eine gute Passung $RMSEA < 0,08$, $CFI > 0,95$ und $SRMR < 0,08$ festgelegt wurden [33]. Modifikationsindizes wurden zur weiteren Untersuchung des Modells herangezogen. Alle Analysen wurden mit der R-Software 4.0.2 und den Paketen dply (1.0.3), tidyr (1.1.2), nFactors (2.4.1) sowie lavaan (0.6-8) durchgeführt.

3 Ergebnisse

Publikation 1: Das ICHOM Standard Set for Pregnancy and Childbirth – Translation and Linguistic Adaption for Germany

Der Großteil der nicht auf Deutsch verfügbaren Items des ICHOM Sets Schwangerschaft und Geburt konnte ohne Probleme übersetzt werden. Zu Abweichungen in der 1:1 Übersetzung kam es vor allem in den Antwortkategorien des BSES_SF und einem Item des BSS-R.

So zeigte sich im Austausch mit den Fachexpertinnen zum BSS-R, dass die Aussage „I came through childbirth virtually unscathed“ übersetzt zu: „Ich habe die Geburt unbeschadet überstanden“ von den Frauen sehr unterschiedlich im Sinne einer körperlichen oder psychischen Unversehrtheit interpretiert werden könnte. Eine Recherche zeigte, dass dies in der sprachlichen Validierung bei der italienischen Übersetzung ebenfalls diskutiert wurde. [34] Im Ergebnis wurde das Wort „unbeschadet“ auf „ohne körperliche oder psychische Komplikationen“ geändert.

Der BSES_EF befasst sich mit dem Stillen und stellt im Original den Aussagen die mit „I can...“ beginnen die Antwortkategorien von „Not at all confident“ bis Very „confident“ gegenüber. Eine deutsche Übersetzung von "confident" kann mit den Begriffen "zuversichtlich", "sicher" und "selbstsicher/selbstbewusst" erfolgen. Es zeigt sich, dass eine direkte Übersetzung der Items nicht immer gut mit den Antwortkategorien zusammenpasste. Diese Erfahrung wurde bereits bei der Übersetzung in Niederländisch gemacht. Nach Beratung mit den Entwickler:innen des Instrumentes und dem Übersetzerteam aus Holland wurde beschlossen, den Zusatz "Ich bin zuversichtlich, dass" am Anfang jedes Items einzufügen. In der Antwortkategorie wurde sich dann ebenfalls für „Zuversichtlich“ als Übersetzung von „confident“ entschieden“.

Im Rahmen der sprachlichen Validierung wurden acht Anpassungen an den Übersetzungen durchgeführt. So zeigte sich das im Kontext von Inkontinenz der fachliche Ausdruck „ungewollte Flatulenzen“ von den Frauen nicht verstanden wurde. Auch der im Jorge Wexner Inkontinenz Score verwendete Begriff „Abgang von Winden“ wurde nicht richtig erfasst. Als Alternative ist häufig der Begriff der Blähungen vorgeschlagen worden.

Als medizinisch präzise und allgemein verständlich wurde im Ergebnis die Formulierung „Abgang von Luft“ gewählt.

Ein zweiter Punkt aus der sprachlichen Validierung ergab sich zu dem Item Schmerzen beim Geschlechtsverkehr. Hier zeigte sich, dass eine vorangestellte Frage zur derzeitigen Aktivität bzw. Zufriedenheit mit dem Sexualleben fehlte. Denn nur bei bestehendem Sexualleben kann auch die Frage nach „Beeinflussung des Sexuallebens durch Schmerz“ beantwortet werden. Ein entsprechendes Item wurde entwickelt und in das Set aufgenommen.

Publikation 2: PROMIS-29 and EORTC QLQ-C30 – an empirical investigation towards a common conception of health

Inhaltsanalyse

Die Gegenüberstellung des Inhalts beider Instrumente ergab eine vergleichbare Konzeptualisierung von Gesundheit sowohl auf der Ebene der Domänen als auch auf der Ebene der Items. Jede der sieben Domänen des PROMIS-29 konnte eine Domäne des QLQ-C30 zugeordnet werden. Ein wesentlicher Unterschied zwischen den beiden Instrumenten besteht in der Anzahl der verwendeten Items zur Messung jeder Domäne. Während im PROMIS-29 vier Items pro Domäne verwendet werden, ist die Anzahl der Items im QLQ-C30 pro Dimension unterschiedlich. Beispielsweise wird Schlaf mit einem Item und Erschöpfung mit drei Items gemessen.

Neben der Vergleichbarkeit des Inhalts zeigt auch die Formulierung in den einzelnen Items weitere Ähnlichkeiten zwischen den beiden Instrumenten. Ein wesentlicher Unterschied besteht insbesondere im Messspektrum der Items. Betrachtet man zum Beispiel die Domäne körperliche Funktion, so umfasst das QLQ-C30 eine breitere Palette an Aktivitäten, einschließlich sehr grundlegender Aufgaben (z. B. Essen, Ankleiden), einfacher Aufgaben (z. B. kurzer Spaziergang) und anspruchsvollerer Aufgaben (z. B. langer Spaziergang), während das PROMIS-29 keine grundlegenden Aufgaben abdeckt. Auf der anderen Seite umfasst die SL-Domäne des PROMIS-29 ein breiteres Spektrum der Schlafqualität im Vergleich zu dem einem Item des QLQ-C30.

Statistische Analysen

In der Kohorte konnten keine signifikanten Abweichungen in Bezug auf die durchschnittlichen T-Werte der entsprechenden Domänen beider Messinstrumente beobachtet werden. Die Abweichungen zeigen nur geringfügige Unterschiede (<5) zueinander auf.

Die EFA ergab sechs Eigenwerte größer als 1 (19,95, 5,08, 3,02, 1,85, 1,39, 1,14), was gemäß dem Kaiser-Kriterium auf sechs zu extrahierende Faktoren hinweist. Auch die Parallelanalyse deutete auf sechs Faktoren hin. Die Betrachtung des Scree-Plots legt hingegen nahe, dass drei bis fünf Faktoren vorhanden sind. Daher wurden vier verschiedene EFA-Modelle mit drei bis sechs Faktoren geschätzt.

Das EFA-Modell mit fünf Faktoren erwies sich als am aussagekräftigsten bei guter Interpretierbarkeit. Diese fünf Faktoren konnten wie folgt kategorisiert werden: Körperliche Funktion, Psychische Gesundheit, Fatigue (Erschöpfung), Schmerz und Schlaf. Sie erklärten zusammen 55,5% der Gesamtvarianz. Alle Items beider Instrumente, die sich auf körperliche Funktion, Erschöpfung, Schmerz und Schlaf bezogen, luden auf denselben Faktor. In der Kategorie Psychische Gesundheit laden alle Items des PROMIS-29. Die vier Items des QLQ-C30 verteilen sich in Ihrer Ladung v.a auf die Kategorien Körperliche Funktion, Psychische Gesundheit und Erschöpfung.

Bei der Modellierung der CFA wurde sich an den Ergebnissen der Inhaltsanalyse (sieben Domänen) und EFA (fünf Domänen) orientiert. So wurde in das Modell Fatigue als Faktor aufgenommen und basierend auf den Ergebnissen der EFA Angst und Depression als gemeinsame Faktoren (Mental Health) modelliert. Abbildung 2 zeigt das resultierende CFA-Modell mit 6 Domänen, dass eine zufriedenstellende Passung der Daten aufweist, mit $RMSEA = 0,074$, $CFI = 0,968$ und $SRMR = 0,074$. Insgesamt haben alle Items hohe Ladungen ($>0,6$) auf den angenommenen Faktoren und die Korrelationen zwischen den Faktoren entsprechen den Erwartungen.

Insgesamt zeigen die Ergebnisse, dass die grundlegende Konzeption von Gesundheit beim QLQ-C30 und dem PROMIS29 vergleichbar ist und somit eine ausreichende Basis dafür besteht Score-Transformationsalgorithmen für diese beiden Instrumente zu entwickeln.

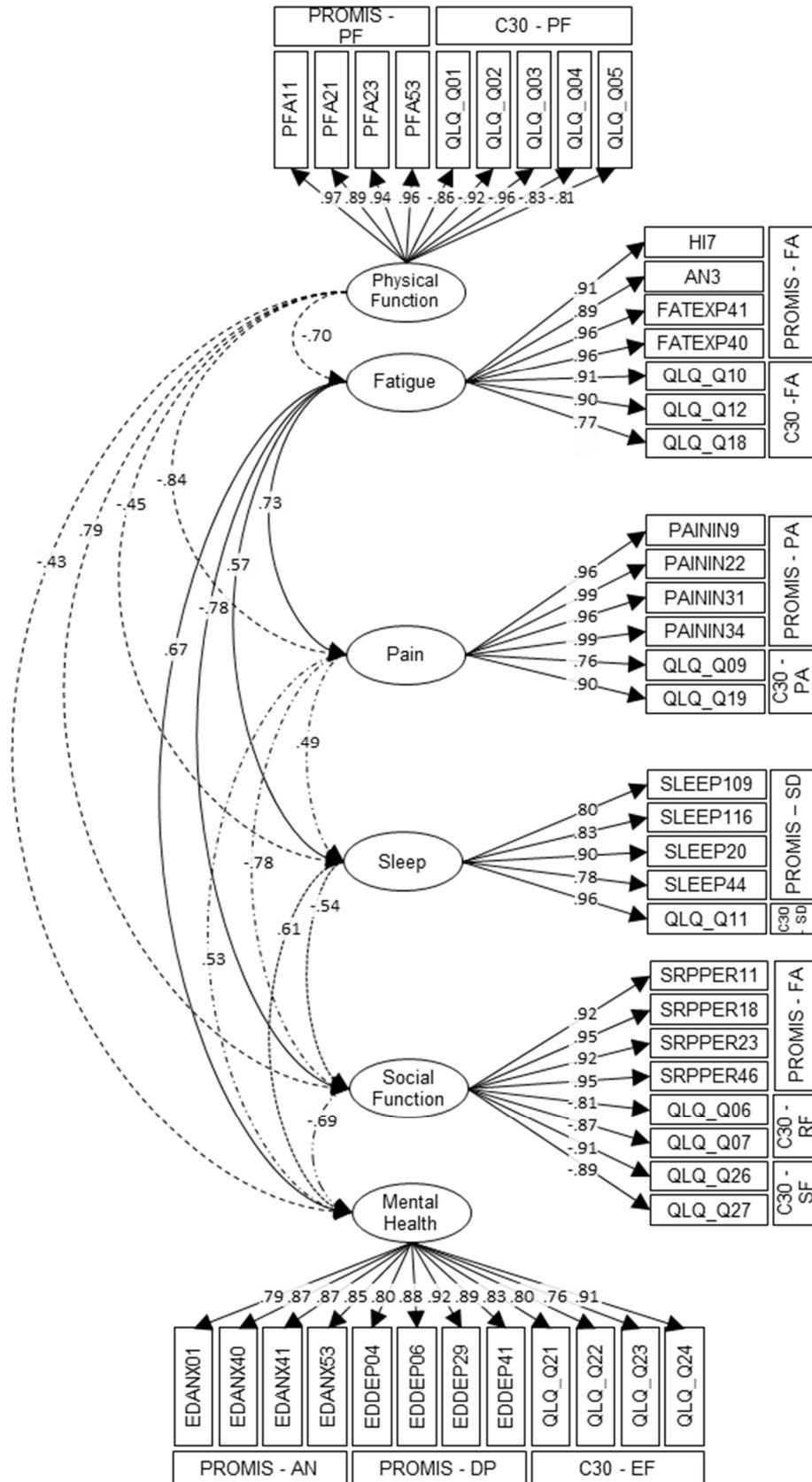


Abbildung 2: CFA-Modell mit 6 Domänen und Ladungen [35]

Anmerkung: AN = Anxiety; DP = Depression; EM = Emotional Function; FA = Fatigue; PA = Pain; PF = Physical Function; RF = Role Function; SRAA = Social Roles And Activities; SD = Sleep Disturbance; SF = Social Function.

4 Diskussion und Implikationen

Zusammenfassung und kritische Würdigung

Eine wesentliche Herausforderung im Kontext der Nutzung patientenberichteter Gesundheitsdaten stellt die Vergleichbarkeit von PRO-Daten dar, die aus unterschiedlichen Kliniken, Ländern, Studien etc. stammen. In dieser Arbeit wurden zwei Wege zur Standardisierung von PRO Erhebungen näher untersucht. Zum einen wurde das ICHOM Standardset für Schwangerschaft und Geburt übersetzt und sprachlich validiert. Somit kann ein international konzertiertes Set und die enthaltenen PROMs nun auch in Deutschland eingesetzt werden. Zum anderen wurden der PROMIS-29 und QLQ-C30 sowohl inhaltlich deskriptiv als auch statistisch hinsichtlich der zugrundeliegenden Konstrukte verglichen. Das Ergebnis zeigt eine hohe Übereinstimmung und liefert die Basis die einzelnen Domänen beider Instrumente über einen Transformationsalgorithmus oder der Kalibrierung auf eine einheitliche Metrik direkt vergleichbar zu machen.

Im Rahmen der Übersetzung des ICHOM Standardsets wurde mit der FACIT Methode eine etablierte Übersetzungsmethode verwendet, um die Gleichwertigkeit von Bedeutung und Messeigenschaften zwischen den Sprachversionen sicherzustellen. Im Zuge des Cognitive Debriefing zeigte sich, dass die Themen des Sets wie z.B. Geburtserleben, Stillen oder Inkontinenz sehr relevant aber auch schambehaftet für die Frauen sind. Aktuelle Studien untermauern die empfundene Relevanz. Beispielsweise zeigte sich, dass nicht allein der empfundene Schmerz mit der Bewertung des Geburtserlebnisses korreliert, sondern Faktoren wie die Beteiligung an Entscheidungen im Geburtsverlauf als auch das Maß der Unterstützung durch das medizinische Personal eine viel wichtigere Rolle bei der Bewertung spielen.[36] Es zeigt sich weiterhin, dass die gesundheitsbezogene Lebensqualität so wie das subjektive Gesundheitsempfinden wesentlich vom Geburtserleben beeinflusst werden [37] und sich beide Faktoren auf die Entwicklung der Mutter-Kind-Beziehung auswirken [38]. Die Implementierung des Sets oder einzelner Instrumente bietet daher eine große Chance zu einer verbesserten Arzt-Patientin-Kommunikation, aber vor allem für eine systematische, patientenzentrierte Bewertung und Verbesserung der Versorgungsqualität in der Geburtsmedizin. Limitierend ist der Arbeit anzumerken, dass zwar eine sprachliche jedoch keine quantitative Validierung durchgeführt wurde, mit der auch die Messeigenschaften der übersetzten Instrumente überprüft wurde.[23]

Zwischen den Instrumenten PROMIS-29 und QLQ-C30 zeigte sich nicht nur eine große inhaltliche Überschneidung auf Domain- und Itemlevel, sondern auch eine empirische in Bezug auf die Korrelationskoeffizienten und das gemessene Outcome (T-Werten). Wobei anzumerken ist, dass beide Instrumente zwar die gleiche Metrik nutzen, diese jedoch an unterschiedlichen Populationen kalkuliert wurden. De facto werden daher zwei verschiedene Werte berechnet. Basierend auf den Analysen konnte mittels EFA ein 5-Faktorenmodell entwickelt werden. Die Übereinstimmung der Modelle PROMIS-29 und QLQ-C30 in der CFA deutet darauf hin, dass beide Instrumente ein ähnliches Konstrukt „allgemeinen Gesundheit“ erfassen. Eine Stärke der Arbeit ist das zugrundeliegende Datenset mit 1.500 Patient:innen. Da für die Arbeit jedoch lediglich Daten von Brustkrebspatientinnen genutzt wurden, lassen sich die Ergebnisse nicht zwingend auf Männer und andere Krankheiten übertragen. Eine weitere Limitation ergibt sich aus der Anzahl der verwendeten Items pro Domain, was zu einer geringeren Messgenauigkeit führt.[35]

Implikationen für Praxis und weitere Forschung

Aufgrund der noch fehlenden Übersetzungstabellen für einen Großteil der Instrumente, scheint eine Standardisierung über einheitlich definierte Sets der derzeit einfachste Weg. Gerade für Kliniker:innen erleichtert die Verfügbarkeit von Standardsets die Auswahl von PROMs. Neben großen Organisationen wie ICHOM und COMET, werden Standardsets auch durch nationale und internationale Fachgesellschaften oder Qualitätsverbände definiert. Dies führt zwar insgesamt zu einer Reduktion der bisherigen Diversität, bedeutet jedoch nicht zwingend, dass nun Ergebnisse miteinander verglichen werden können, da diese Sets meist unterschiedliche Instrumente enthalten. Weitere Nachteile bei dem Einsatz von Standardsets ist die Länge dieser, die Abhängigkeiten bei Lizenzänderungen und mangelnde Flexibilität v.a. im Hinblick auf die Einbindung neuer Instrumente oder Messmethoden.

Die Festlegung einer einheitlichen Metrik stellt einen zeitgemäßen und methodisch anspruchsvollen Ansatz zur Standardisierung dar. Er ermöglicht größte Flexibilität und einen niedrigen Grad an Informationsverlust, da auch ältere Datenerhebungen auf der T-Metrik abgebildet bzw. in einen T-Score umgerechnet werden können. Für die Auswahl

eines PROM hat dies zwei Implikationen: So kann der Anwender frei wählen, welches Instrument verwendet wird, ohne einen Verlust der Vergleichbarkeit riskieren zu müssen. Zum anderen kann die Entscheidung jederzeit geändert und auf ein anderes Instrument gewechselt werden ohne einen Informationsverlust bzgl. der vorangegangenen Erhebung einzugehen. Ein weiterer Vorteil ist die Nutzung von Computer adaptiven Tests, hierbei wird die Itemauswahl basierend auf dem individuellen Antwortverhalten abgefragt und nicht zwingend das gesamte Instrument. Auf Grund des vergleichsweise hohen Aufwands bei der Auswahl eines bereits kalibrierten Instrumentes sowie der limitierten Verfügbarkeit solcher, ist der Ansatz derzeit nur bedingt in der Breite umsetzbar. Weitere Forschung und vor allem Kalibrierung von etablierten Instrumenten an die T-Metrik ist notwendig, damit sich der Ansatz in der Praxis weiter durchsetzt. Hierzu zählt auch die Übersetzung von Items und Instrumenten, um diese in möglichst vielen Ländern einsetzbar zu machen.

Anders als in anderen europäischen Ländern (u.a. Dänemark, Niederlande) findet sich in Deutschland keine politische Strategie zum Thema PROM-Einsatz, was eine Standardisierung in jeglicher Hinsicht deutlich erschwert.[39] Einzig in der Stellungnahme der Regierungskommission mit dem Thema „Weiterentwicklung der Qualitätssicherung, des Qualitäts- und des klinischen Risikomanagements. Mehr Qualität – weniger Bürokratie“ ist zu sehen, dass PROs als Bewertungs- und Steuerungsinstrument auch auf politischer Ebene als relevant erachtet werden. So wird in dem Bericht gefordert, dass PROMs für alle Krankenhäuser verbindlich erhoben und Teil der qualitätsabhängigen Vergütung werden. Neben einzelnen Krankenhäusern oder medizinischen Fachgesellschaften gibt es eine größere Initiative, welche die PRO-Erhebung vorantreibt. Die Initiative Qualitätsmedizin (IQM) möchte Ihre bisherige Qualitätserhebung um PRO-Daten erweitern. Über 500 Mitgliedskliniken sollen zukünftig PROMs erheben.[40] Dafür werden Standardsets für die häufigsten Krankheitsbilder entwickelt. Diese werden aus einem generischen, krankheitsunabhängigen Instrument bestehen, welches bereits auf der T-Metrik verankert ist, sowie aus Krankheits- oder therapiespezifischen Instrumenten. Dieses Vorgehen zeigt, dass eine Kombination beider Standardisierungsansätze derzeit am praktikabelsten ist.

Literaturverzeichnis

1. Regierungskommission. (2023). Siebente Stellungnahme und Empfehlung der Regierungskommission für eine moderne und bedarfsgerechte Krankenhausversorgung: Weiterentwicklung der Qualitätssicherung, des Qualitäts und des klinischen Risikomanagements (QS, QM und kRM) - Mehr Qualität – weniger Bürokratie. (pp. 43). online: Regierungskommission für eine moderne und bedarfsgerechte Krankenhausversorgung.
2. Porter, M. E., Michael Porter, E. O. T., & Teisberg, E. O. (2006). Redefining Health Care: Creating Value-based Competition on Results: Harvard Business School Press
3. Hartmann, C., Obbarius, A., Haneke, H., Klapproth, C. P., & Rose, M. (2021). Value-based Healthcare – Gesundheitsversorgung neu denken. OP-Management up2date, 1(01), 87-105
4. U.S. Food and Drug Administration. (2009). Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Guidance for Industry. In U.S. Food and Drug Administration (Ed.).
5. NHS. (2023). A guide to PROMs methodology. online: National Health Service.
6. Nilsson, E., Orwelius, L., & Kristenson, M. (2016). Patient-reported outcomes in the Swedish National Quality Registers. J Intern Med, 279(2), 141-153,doi:10.1111/joim.12409.
7. Ernstsson, O. (2021). Making use of patient-reported outcome measures in health care : the case of EQ-5D in the Swedish national quality registries. Karolinska Institutet.
8. The Danish Health Data Authority. (2018). Digital Health Strategy 2018-2022. online.
9. Bertelsmann Stiftung. (2023). Patient-Reported Outcomes - Mit patientenberichteten Daten zu einer besseren Versorgungsqualität. online
10. Bergner, M., Bobbitt, R. A., Carter, W. B., & Gilson, B. S. (1981). The Sickness Impact Profile: Development and Final Revision of a Health Status Measure. Medical Care, 19(8), 787-805
11. eProvide by Mapi Research Trust. Zugriff: 04.10.2023, from <https://eprovide.mapi-trust.org/advanced-search?search=Depression>
12. Deyo, R. A., Battie, M., Beurskens, A. J., Bombardier, C., Croft, P., Koes, B., Malmivaara, A., Roland, M., Von Korff, M., & Waddell, G. (1998). Outcome measures for low back pain research. A proposal for standardized use. Spine (Phila Pa 1976), 23(18), 2003-2013,doi:10.1097/00007632-199809150-00018.
13. Bhashyam, A. R., van der Vliet, Q. M. J., Ochen, Y., Heng, M., Leenen, L. P. H., Hietbrink, F., & Houwert, R. M. (2019). Injury-related variation in patient-reported outcome after musculoskeletal trauma: a systematic review. Eur J Trauma Emerg Surg,doi:10.1007/s00068-019-01261-8.
14. Maspero, M., & Hull, T. (2023). Patient-Reported Outcomes in Colorectal Surgery. Clin Colon Rectal Surg, 36(4), 240-251,doi:10.1055/s-0043-1761607.
15. Porter, M. E., Larsson, S., & Lee, T. H. (2016). Standardizing Patient Outcomes Measurement. New England Journal of Medicine, 374(6), 504-506,doi:10.1056/NEJMp1511701.
16. ICHOM. Zugriff: 04.10.2023, from <https://www.ichom.org/aboutus/>
17. Wahl, I., Löwe, B., Bjorner, J. B., Fischer, F., Langs, G., Voderholzer, U., Aita, S. A., Bergemann, N., Brähler, E., & Rose, M. (2014). Standardization of depression

- measurement: a common metric was developed for 11 self-report depression measures. *J Clin Epidemiol*, 67(1), 73-86,doi:10.1016/j.jclinepi.2013.04.019.
18. Fried, E. I. E. d. B. (2021). From mandating common measures to mandating common metrics: a plea to harmonize measurement results. *PsyArXiv*,doi:10.31234/osf.io/m4qzb.
 19. U.S. Food and Drug Administration. (2006). Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes*, 4, 79,doi:10.1186/1477-7525-4-79.
 20. WHO. (1978). Erklärung von Alma-Ata (1978), Konferenz über primäre Gesundheitsversorgung: WHO - Weltgesundheitsorganisation Europa.
 21. Obbarius, A., Fischer, K. I., Fischer, F., Liegl, G., Obbarius, N., Nolte, S., & Rose, M. (2018). Empirische Erfassung subjektiver Gesundheitsmerkmale am Beispiel der gesundheitsbezogenen Lebensqualität. *Psychother Psychosom Med Psychol*, 68(12), 534-547,doi:10.1055/a-0764-4691.
 22. Nijagal, M. A., Wissig, S., Stowell, C., Olson, E., Amer-Wahlin, I., Bonsel, G., Brooks, A., Coleman, M., Devi Karalasingam, S., Duffy, J. M. N., Flanagan, T., Gebhardt, S., Greene, M. E., Groenendaal, F., JR, R. J., Kowaliw, T., Lamain-de-Ruiter, M., Main, E., Owens, M., Petersen, R., Reiss, I., Sakala, C., Speciale, A. M., Thompson, R., Okunade, O., & Franx, A. (2018). Standardized outcome measures for pregnancy and childbirth, an ICHOM proposal. *BMC Health Serv Res*, 18(1), 953,doi:10.1186/s12913-018-3732-3.
 23. Hartmann, C., Rose, M., Weichert, A., & Weißhaupt, K. (2022). The ICHOM Standard Set for Pregnancy and Childbirth - Translation and Linguistic Adaptation for Germany. *Geburtshilfe Frauenheilkd*, 82(7), 747-754,doi:10.1055/a-1666-0429.
 24. Eremenco, S. L., Cella, D., & Arnold, B. J. (2005). A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Eval Health Prof*, 28(2), 212-232,doi:10.1177/0163278705275342.
 25. Willis, G. (2015). *Analysis of the Cognitive Interview in Questionnaire Design: Oxford University Press; Illustrated Edition*
 26. Konrad, K. L. D. I. M., Günter, & Mruck, K. H. H. Q. F. i. d. P. W. S. V.-. (2010). Lautes Denken. In G. May & K. Mruck (Eds.), *Handbuch Qualitative Forschung in der Psychologie*. (pp. 476-490). Wiesbaden: Springer Verlag.
 27. Cella, D., Choi, S. W., Condon, D. M., Schalet, B., Hays, R. D., Rothrock, N. E., Yount, S., Cook, K. F., Gershon, R. C., Amtmann, D., DeWalt, D. A., Pilkonis, P. A., Stone, A. A., Weinfurt, K., & Reeve, B. B. (2019). PROMIS(®) Adult Health Profiles: Efficient Short-Form Measures of Seven Health Domains. *Value Health*, 22(5), 537-544,doi:10.1016/j.jval.2019.02.004.
 28. Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., Filiberti, A., Flechtner, H., Fleishman, S. B., de Haes, J. C., & et al. (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst*, 85(5), 365-376,doi:10.1093/jnci/85.5.365.
 29. EORTC. CORE QUESTIONNAIRES Zugriff: 05.07.2021, from <https://qol.eortc.org/core/>
 30. Liegl, G., Petersen, M. A., Groenvold, M., Aaronson, N. K., Costantini, A., Fayers, P. M., Holzner, B., Johnson, C. D., Kemmler, G., Tomaszewski, K. A., Waldmann, A., Young, T. E., Rose, M., & Nolte, S. (2019). Establishing the European Norm for the health-related quality of life domains of the computer-

- adaptive test EORTC CAT Core. *European Journal of Cancer*, 107, 133-141,doi:10.1016/j.ejca.2018.11.023.
31. Fischer, F., Gibbons, C., Coste, J., Valderas, J. M., Rose, M., & Leplège, A. (2018). Measurement invariance and general population reference values of the PROMIS Profile 29 in the UK, France, and Germany. *Quality of Life Research*, 27(4), 999-1014,doi:10.1007/s11136-018-1785-8.
 32. Wechsler, S. (1997). *Statistics at Square One*. Ninth Edition, revised by M. J. Campbell, T. D. V. Swinscow, BMJ Publ. Group, London, 1996. No. of pages: 140. Price: £11. ISBN 0-7279-0916-9. *Statistics in Medicine*, 16(22), 2629-2630,doi:10.1002/(SICI)1097-0258(19971130)16:22<2629::AID-SIM698>3.0.CO;2-Z.
 33. Brown, T. A. (2006). *Confirmatory factor analysis for applied research*: New York : Guilford Press, [2006] ©2006
 34. Nespoli, A., Colciago, E., Pedroni, S., Perego, S., & Fumagalli, S. (2018). The Birth Satisfaction Scale-Revised (BSS-R): process of translation and adaptation in an Italian context. *Ann Ist Super Sanita*, 54(4), 340-347,doi:10.4415/ann_18_04_11.
 35. Hartmann, C., Fischer, F., Klapproth, C. P., Röhle, R., Rose, M., & Karsten, M. M. (2023). PROMIS-29 and EORTC QLQ-C30: an empirical investigation towards a common conception of health. *Qual Life Res*, 32(3), 749-758,doi:10.1007/s11136-022-03324-7.
 36. Hodnett, E. D. (2002). Pain and women's satisfaction with the experience of childbirth: A systematic review. *American Journal of Obstetrics and Gynecology*, 186(5, Supplement), S160-S172,doi:10.1016/S0002-9378(02)70189-0.
 37. Schäfers, R. (2015). Einfluss des Geburtserlebens auf die subjektive Gesundheitseinschätzung von Frauen. *GMS Z Hebammenwiss*, 2:Doc02,doi:doi: 10.3205/zhwi000005.
 38. Rossen, L., Mattick, R. P., Wilson, J., Clare, P. J., Burns, L., Allsop, S., Elliott, E. J., Jacobs, S., Olsson, C. A., & Hutchinson, D. (2019). Mother-Infant Bonding and Emotional Availability at 12-Months of Age: The Role of Early Postnatal Bonding, Maternal Substance Use and Mental Health. *Matern Child Health J*, 23(12), 1686-1698,doi:10.1007/s10995-019-02809-1.
 39. Viktoria Steinbeck, S.-C. E., Dr. Christoph Pross. (2021). *Patient-Reported Outcome Measures (PROMs): ein internationaler Vergleich*. online: Bertelsmann Stiftung,doi: 10.11586/2021053.
 40. initiative Qualitätsmedizin. (2022). PROM (Patient-Reported Outcome Measures). Zugriff am: 24.01.2024, from <https://www.initiative-qualitaetsmedizin.de/projekte-kampagnen>

Eidesstattliche Versicherung

„Ich, Claudia Hartmann, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema:

„Vergleichbarkeit von Patient-Reported Outcome Measures: Zwei Ansätze zur Standardisierung“

‘Comparability of patient-reported outcome measures: Two approaches to standardization‘

selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilserklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Erstbetreuer/in, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; www.icmje.org) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum

Unterschrift

Anteilserklärung an den erfolgten Publikationen

Claudia Hartmann hatte folgenden Anteil an den folgenden Publikationen:

Publikation 1: Hartmann C, Rose M, Weichert A, Weißhaupt K. *Das ICHOM-Standardset für Schwangerschaft und Geburt – Übersetzung und sprachliche Anpassungen für Deutschland*, GebFra, 2022

Beitrag im Einzelnen:

Konzeption und Organisation: Die Promovendin konzipierte den Ablauf der Studie, organisierte Ethikantrag, Übersetzungsleistungen und koordinierte die Zusammenarbeit mit den Teammitgliedern.

methodische Durchführung: Das Cognitive Debriefing mit den Schwangeren und Wöchnerinnen führte die Promovendin zusammen mit der Letztautorin durch. Der Ablauf der Interviews sowie notwendige Unterlagen (u.a. Studieneinwilligung, -aufklärung) wurden von der Promovendin erstellt. Die Konsolidierung der Übersetzungen erfolgte zusammen mit der Letztautorin.

Erstellung des Manuskriptes: Ein erster vollständiger Entwurf des Manuskripts wurde von der Promovendin eigenständig verfasst. Die Tabelle 1, Abbildung 2 und Tabelle 4 wurde von der Promovendin eigenständig erstellt. Das Manuskript wurde von der Promovendin selbstständig eingereicht. Anmerkungen im Revisionsprozess gab es keine.

Publikation 2: Hartmann C, Fischer F, Klapproth CP, Röhle R, Rose M, Karsten MM. *PROMIS-29 and EORTC QLQ-C30 – an empirical investigation towards a common conception of health*, Quality of Life Research, 2023

Beitrag im Einzelnen:

Konzeption und Organisation: Die Promovendin konzipierte den Ablauf der Studie zusammen mit dem Letztautor und koordinierte die Zusammenarbeit mit den Ko-Autoren.

methodischen Durchführung und Datenanalyse: Die Promovendin führte die Auswertung für die deskriptive Statistik sowie die Berechnung der EFA und CFA selbstständig durch und unternahm eine erste Analyse der Ergebnisse. Die finale Analyse und Auswertung der EFA und CFA fand gemeinsam mit den Ko-Autoren statt.

Erstellung des Manuskriptes: Ein erster Entwurf des Manuskripts inkl. Literatur wurde von der Promovendin eigenständig verfasst. Zusammen mit den Ko-Autoren erfolgte die Finalisierung des Artikels. Die Promovendin erstellte alle Tabellen und Grafiken der Publikation. Das Manuskript wurde von der Promovendin selbstständig eingereicht und im Rahmen des Revisionsprozesses verfasste die Promovendin eigenständig die erste Version der Antworten an die Reviewenden und diskutierte diese kritisch mit den Ko-Autoren.

Unterschrift, Datum und Stempel des
erstbetreuenden Hochschullehrers

Unterschrift der Doktorandin

Auszug aus der Journal Summary List

Publikation 1: *Das ICHOM-Standardset für Schwangerschaft und Geburt – Übersetzung und sprachliche Anpassungen für Deutschland* GebFra, 2022

Journal Data Filtered By: **Selected JCR Year: 2022** Selected Editions: SCIE,SSCI
 Selected Categories: **“OBSTETRICS and GYNECOLOGY”** Selected Category
 Scheme: WoS
Gesamtanzahl: 86 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfaktor
35	PRENATAL DIAGNOSIS	6,975	3.0	0.00619
36	CONTRACEPTION	7,236	2.9	0.00943
37	Journal of Perinatology	9,444	2.9	0.01119
38	Reproductive Sciences	6,534	2.9	0.00630
39	CLIMACTERIC	3,568	2.8	0.00338
40	INTERNATIONAL JOURNAL OF OBSTETRIC ANESTHESIA	1,972	2.8	0.00193
41	PAEDIATRIC AND PERINATAL EPIDEMIOLOGY	3,933	2.8	0.00375
42	Breastfeeding Medicine	3,043	2.7	0.00415
43	GEBURTSHILFE UND FRAUENHEILKUNDE	1,577	2.7	0.00153

Publikation 2: *PROMIS-29 and EORTC QLQ-C30 – an empirical investigation towards a common conception of health*, Quality of Life Research, 2023

Journal Data Filtered By: **Selected JCR Year: 2022** Selected Editions: SCIE,SSCI
 Selected Categories: **“HEALTH POLICY and SERVICES”**
 Selected Category Scheme: WoS
Gesamtanzahl: 87 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfaktor
21	American Journal of Health Economics	382	3.7	0.00190
22	Applied Health Economics and Health Policy	2,159	3.6	0.00364
23	Health and Quality of Life Outcomes	14,636	3.6	0.01232
24	Health Care Management Science	1,915	3.6	0.00165
25	Health Sociology Review	876	3.6	0.00106
26	MEDICAL DECISION MAKING	6,872	3.6	0.00582
27	Patient-Patient Centered Outcomes Research	2,282	3.6	0.00396
28	International Journal of Health Policy and Management	3,362	3.5	0.00675
29	JOURNAL OF HEALTH ECONOMICS	8,959	3.5	0.01053
30	QUALITY OF LIFE RESEARCH	21,482	3.5	0.01890

Druckexemplar(e) der Publikation(en)

GebFra Science | Original Article



The ICHOM Standard Set for Pregnancy and Childbirth – Translation and Linguistic Adaptation for Germany

Das ICHOM-Standardset für Schwangerschaft und Geburt – Übersetzung und sprachliche Anpassung für Deutschland



Authors

Claudia Hartmann¹, Matthias Rose¹, Alexander Weichert², Karen Weißhaupt²

Affiliations

- 1 Medizinische Klinik m. S. Psychosomatik, Charité – Universitätsmedizin Berlin, Berlin, Germany
- 2 Klinik für Geburtsmedizin, Charité – Universitätsmedizin Berlin, Berlin, Germany

Key words

subjective perception of health, patient-reported outcomes, breastfeeding, early bonding, birth experience

Schlüsselwörter

subjektives Gesunderleben, Patient-Reported Outcomes, Stillen, frühe Bindungsbildung, Geburtserleben

received 3.9.2021
accepted 9.10.2021

Bibliography

Geburtsh Frauenheilk 2022; 82: 747–754
DOI 10.1055/a-1666-0429
ISSN 0016-5751

© 2022. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Georg Thieme Verlag KG, Rüdigerstraße 14,
70469 Stuttgart, Germany

Correspondence

Claudia Hartmann, Dipl. Kffr.
Charité – Universitätsmedizin Berlin,
Medizinische Klinik m. S. Psychosomatik
Charitéplatz 1, 10117 Berlin, Germany
claudia.hartmann@charite.de



Deutsche Version unter:
<https://doi.org/10.1055/a-1666-0429>
Supplementary material is available under
<https://doi.org/10.1055/a-1666-0429>

ABSTRACT

Introduction The gathering of health and quality-of-life issues that matter most to patients over so-called patient-reported outcomes (PROs) is a key aspect of patient-centered healthcare. In obstetrics, this approach has the potential to expand the current understanding of what quality entails by including patient perspectives. The International Consortium for Health Outcome Measures (ICHOM), founded in 2012, is a global organization which aims to standardise the collection of PROs and make the results comparable worldwide. A PRO Set for obstetrics, “Pregnancy and Childbirth”, was published in 2018. The aim of our work was to translate the instruments of this set that are not yet available in German into German.

Methods The instruments were translated from English into German using the Functional Assessment of Chronic Illness Therapy (FACIT) translation method. The translated instruments consisted of the Breastfeeding Self-Efficacy Scale – Short Form (BSES-SF), the Mother-to-Infant Bonding Scale (MIBS), and the Birth Satisfaction Scale – Revised (BSS-R) as well as a number of individual questions (items).

Results The final version of the German translations were found to be easily comprehensible by the target group. During the cognitive debriefing, it became clear that pregnant women and women who had just given birth often used a number of medically incorrect terms to refer to their symptoms or complaints. In the translations great care was taken to ensure that the wording was as close as possible to the general usage of the language while at the same time the terminology was medically correct. To achieve a precise but comprehensible translation, the response structure of the BSES-SF also had to be adapted.

Conclusions The instruments of the ICHOM Standard Set “Pregnancy and Childbirth”, which were not previously available in German, were successfully translated into German. This meant dealing with a few challenges such as adapting questions or response structures. The cultural and linguistic comprehensibility of the German translations were confirmed during the subsequent cognitive debriefing. The translations offer the possibility of implementing the complete ICHOM

Set for Pregnancy and Childbirth. This would provide an opportunity to expand the existing understanding of quality by including the subjective experience of women during and after childbirth and, in future, to compare outcomes with those of other hospitals across the world.

ZUSAMMENFASSUNG

Einleitung Die Erhebung von patient*innenberichteter Gesundheit und Lebensqualität über sogenannte Patient-Reported Outcomes (PROs) – sind ein zentraler Bestandteil patient*innenzentrierter Gesundheitsversorgung. Auch in der Geburtsmedizin haben sie das Potenzial, das aktuelle Qualitätsverständnis um die Patientinnenperspektive zu erweitern. Mit dem International Consortium for Health Outcome Measures (ICHOM) gibt es seit 2012 eine globale Organisation, welche die Erhebung von PROs standardisieren und Ergebnisse weltweit vergleichbar machen möchte. Für den Bereich Geburtsmedizin wurde 2018 das PRO-Set „Pregnancy and Childbirth“ veröffentlicht. Ziel dieser Arbeit ist die Übersetzung der noch nicht im Deutschen verfügbaren Instrumente dieses Sets ins Deutsche.

Methoden Die Übersetzung der Instrumente vom Englischen ins Deutsche erfolgte nach der Functional Assessment of Chronic Illness Therapy-(FACIT-)Übersetzungsmethode. Die übersetzten Instrumente umfassen die Breastfeeding Self-Efficacy Scale – Short Form (BSES-SF), die Mother-to-Infant

Bonding Scale (MIBS), die Birth Satisfaction Scale – Revised (BSS-R) sowie mehrere Einzelfragen (Items).

Ergebnisse Die Endfassung der deutschen Übersetzungen ergaben ein hohes Maß an Verständlichkeit in der Zielgruppe. Bei der Durchführung des Cognitive Debriefings zeigte sich, dass im Sprachgebrauch der Schwangeren und Wöchnerinnen häufig medizinisch inkorrekte Formulierungen für Symptome oder Beschwerden genutzt werden. In den Übersetzungen wurde daher darauf geachtet, die Formulierungen möglichst dem allgemeinen Sprachgebrauch bei gleichzeitig medizinisch korrekter Ausdrucksweise anzupassen. Des Weiteren musste, um eine präzise und verständliche Übersetzung zu erreichen, die Antwortstruktur des BSES-SF adaptiert werden.

Schlussfolgerungen Die bisher nicht in deutscher Sprache verfügbaren Instrumente des ICHOM-Sets „Pregnancy and Childbirth“ konnten mit einigen wenigen Herausforderungen, wie z. B. Anpassung der Frage- oder Antwortstruktur, erfolgreich ins Deutsche übersetzt werden. Die kulturelle und sprachliche Verständlichkeit der deutschen Übersetzung konnte im Rahmen des Cognitive Debriefings bestätigt werden. Die Übersetzung bietet die Möglichkeit, das vollständige ICHOM-Set für Schwangerschaft und Geburt zu implementieren. Daraus ergibt sich die Chance, das bisherige Qualitätsverständnis um das subjektive Erleben der Frauen rund um die Geburt zu erweitern sowie Ergebnisse zukünftig mit Kliniken international zu vergleichen.

Introduction

“A good birth goes beyond having a healthy baby.” [1]. Based on growing scientific knowledge, it has been possible to increasingly standardise the care of pregnant women and the peripartum somatic care of women in recent decades. This is reflected by the increase in both the quality and quantity of international and national guidelines, such as those issued by the German Society of Gynecology and Obstetrics (DGGG) [2]. Currently, procedural and physical data about the course of a birth and the state of health of the mother and the newborn are primarily collected and evaluated to monitor the quality of established guidelines. The focus is first and foremost on ensuring low morbidity and mortality rates. A more in-depth analysis of the data provides additional insight into current trends, e.g., the percentage of cesarean sections, the number of medically induced births, or treatment effects following the use of new or amended methods. Taken together, the data provide some insight into the quality and extent of care provided. But these physical and procedural data do not shed much light on either the birth as such or on how the convalescence process was experienced from the perspective of the women; which medium to long-term impact did the birth have on their body image, their physical and social activities, their continence, their breastfeeding and the relationship between mother and child. Including the patients' perspective is therefore fundamental, particularly in the context of a holistic, future-oriented medicine [3].

In recent years, the standardized collection of patient-reported outcomes (PROs) to evaluate the current state of health and the quality of treatment received from the patients' perspective has become increasingly important in many countries and in all areas of medical care. In the United Kingdom, PROs have been used systematically since 2009 as indicators for the quality of care given to patients who receive hip or knee replacements [4]. In the Netherlands, the use of PROs was stipulated in the government's coalition agreement of 2017, and the Ministry of Health has submitted a concept for the gradual introduction of PROs [5].

PROs include all characteristics or manifestations that relate to a patient's state of health as reported by the patient herself or himself, without these data being altered or interpreted by someone else (in particular, by medically trained staff) [6]. In addition to concrete symptoms, this also encompasses quality-of-life factors, including physical as well as psychological and social factors [7]. PROs are measured with the help of Patient Reported Outcome Measures (PROMs), generally referred to as questionnaires or instruments. For example, the Edinburgh Postnatal Depression Scale (EPDS = PROM) can be used to measure postpartum depression in women (= PRO). A number of instruments are currently available to measure depression, but they use different scales to present their results. In order to enable future exchange and comparison of results collected with different instruments in hospitals worldwide, it is necessary that the instruments are either collected with the same metric (unit) or all use the same questionnaire.

Internationally there are two leading cross-disease initiatives which aim to create a way of standardizing and comparing such patient-related data. One of them is PROMIS (Patient Reported Outcomes Measurement Information System) and the other is the International Consortium for Health Outcomes Measurement. In addition to providing their own instruments to survey disease-independent health factors, PROMIS aims to standardize the evaluation of outcomes and results by establishing a standard set of metrics, with "50" corresponding to the average of the general population, and a standard deviation of "10" (T-scores) [8].

ICHOM is an initiative which compiles sets of questions taken from previously validated instruments for various clinical pictures across a number of medical areas (including interdisciplinary areas). The aim is to use the same instruments to standardize and compare both national and international outcomes. As part of the development process to design an ICHOM Standard Set, an international working group composed of clinicians, researchers and patient representatives define a minimum number and standard of questionnaires and corresponding case-mix factors, which should be administered before the start of treatment and during convalescence. These sets may partly include PROMIS instruments.

In 2018, a 21-member international panel of experts composed of midwives, obstetricians/gynecologists and patient representatives from a total of 8 different countries developed a Standard Set for surveys to be administered during pregnancy and childbirth. The relevant domains (factors), case-mix parameters, questionnaires and time points when the survey instruments should be administered were decided in 11 sessions, using a modified Delphi method to establish consensus. In addition to a systematic literature analysis, this method includes the compilation of expert opinions and interviews with pregnant women and

mothers. Particularly interviews with affected women aim to ensure that patients' perspectives are taken sufficiently into account when evaluating care [9].

Methods

Instruments

► **Table 1** shows the health and quality-of-life factors which are surveyed with the ICHOM Set for Pregnancy and Childbirth. In addition to patient demographics and patient history, the surveys focus on the dimensions "incontinence", "breastfeeding", "change of role", "birth experience", "general health" and "mental health" as well as "pain during sexual intercourse". The eight instruments and 19 individual questions resulted in a list of 82 items. Five of the instruments (31 items in total) and 3 individual questions were already available in German in the form of validated translations [10–14]. The Breastfeeding Self-Efficacy Scale – Short Form (BSES-SF), the Mother-to-Infant Bonding Scale (MIBS), the Birth Satisfaction Scale – Revised (BSS-R) and several individual questions had not yet been translated. The design of the ICHOM Standard Set envisages a total of five different time points when a survey should be administered (► **Fig. 1**), starting with the first prepartum examination before the third month of pregnancy and concluding with a survey carried out at six months postpartum.

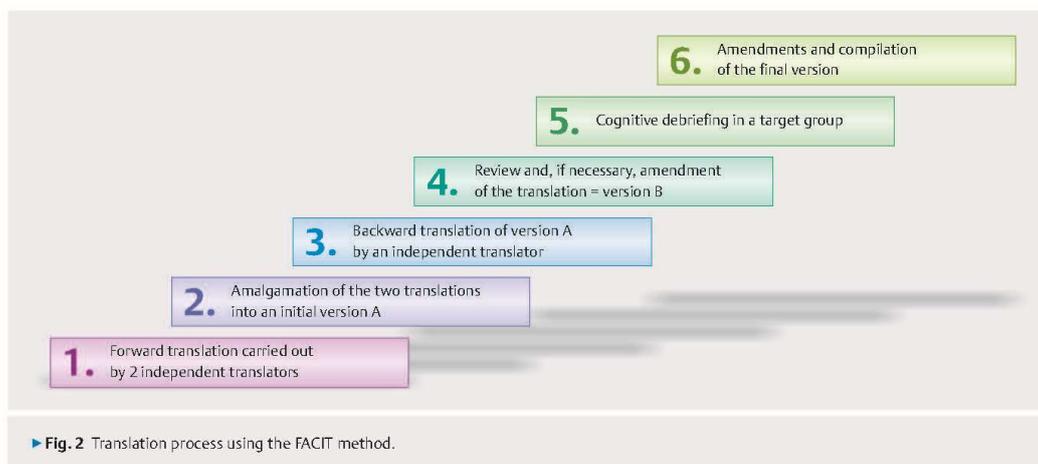
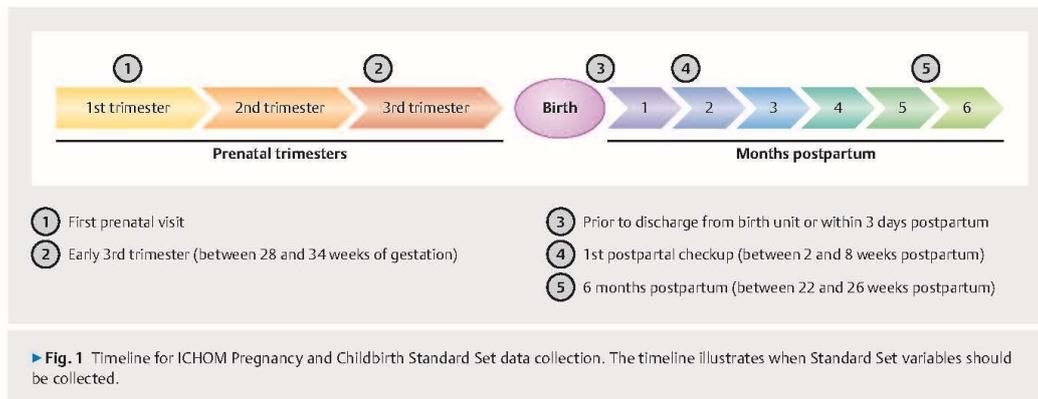
Translation

The three instruments BSES-SF, MIBS and BSS-R, which were not yet available in German (a total of 32 items), and 16 individual questions were translated using the Functional Assessment of Chronic Illness Therapy (FACIT) translation method (► **Fig. 2**). The

► **Table 1** Overview of the ICHOM Standard Set for Pregnancy and Childbirth.

Factors	Instruments and items	German translation	Time points to administer a survey*
Patient demographics and history	Individual questions (8)	✓	1
Incontinence	Jorge-Wexner Incontinence Score or ICIQ-UI-SF	✓ [10] ✓ [11]	1, 2, 4, 5
	+ Individual questions: incontinence screening (1)	–	1, 2, 4, 5
Breastfeeding	Breastfeeding Self-Efficacy Scale – Short Form (BSES-SF)	–	2, 3, 4
	+ individual questions: "successful breastfeeding" (3)	–	3, 4, 5
Role transition	Mother-to-Infant Bonding Scale (MIBS)	–	3, 4
	+ individual question: confidence in role as mother (1)	–	4, 5
Birth experience and satisfaction with care received	Birth Satisfaction Scale – Revised (BSS-R)	–	4
	+ individual questions: satisfaction with care received (5)	–	2, 4, 5
General health and quality of life	PROMIS Global-10	✓ [12]	1, 2, 4, 5
Mental health	PHQ-2	✓ [13]	1, 2, 4, 5
	+ Edinburgh Postnatal Depression Scale (EPDS)	✓ [14]	
Pain with intercourse	PROMIS Sexual Function (single item)	–	1, 4, 5

PHQ = Patient Health Questionnaire, PROMIS = Patient Reported Outcomes Measurement Information System, * cf. ► **Fig. 1**.



FACIT method is a translation process which ensures the semantic and technical equivalence of translations to the original texts [15]. The permission of the respective instrument developers was sought before starting to translate the instruments BSES-SF, MIBS and BSS-R.

In accordance with the FACIT method, the first step consisted of a forward translation into German by two independent translators. Both translations were subsequently merged together to create a first German-language version. In the next step, this version was translated back into English, again by two independent translators. Any discrepancies between the English back translation and the original English text were discussed by the research team and, if considered sufficiently important, the German wording was changed or amended.

Linguistic validation

Cognitive debriefing is a linguistic validation process in which possible problems of comprehension relating to individual items are tested in different target groups [16]. The 48 items requiring validation were divided into two groups according to the time point at which the questions were administered, with 29 items tested in a target group of pregnant women and 20 items tested in a target group of women who have given birth. The question about "Confidence with role as a mother" was validated in both groups, as there were differences in the wording used for the prenatal and postpartum surveys. Because of the large number of items (29 items) requiring validation by the group of pregnant women, this group of items was subdivided. The items were ultimately tested in three different groups:

1. pregnant women = 10 items,
2. pregnant women = 19 items, and
3. new mothers = 20 items.

A total of 15 women were recruited for the validation, with five women in each of the three groups. All of the women were currently receiving care in the Charité, Department of Obstetrics (prepartum care or mother-and-child ward). The participants were requested to read the entire item aloud, reproduce the contents of the item in their own words and subsequently “think aloud” about the item. As they were thinking aloud, they were requested to address anything they stumbled over while reading aloud or which appeared unclear to them [17]. Considerations and comments made during the cognitive debriefing were subsequently discussed and incorporated in the translation.

Results

Translation process

The majority of items could be translated forwards and backwards without difficulty. A direct translation could not be used for one of the items of the BSS-R nor for the response categories of the BSES-SF.

BSS-R

There was no difficulty in translating the question “I came through childbirth virtually unscathed” in the BSS-R literally into German as “*Ich habe die Geburt unbeschadet überstanden*”, and the backward translation of the question was found to closely match the original. However, when the translations were reviewed and merged by obstetricians, it was noted that women interpreted this phrase in a number of very different ways. The same issue had previously been raised by the experts reviewing the Italian translation of this item. They had noted that the word “unscathed” could be understood psychologically, physically or emotionally [18]. Following this analysis, the German word “*unbeschadet*” was replaced by the phrase “*ohne körperliche oder psychische Komplikationen*” (which translates literally as “without physical or mental complications”).

BSES-SF

The possible response categories were given as a scale which ranged from “Not at all confident” to “Very confident”, although the translation of “confident” into German is ambiguous as the word can also be translated as self-assured/self-confident, certain or optimistic. After talking to the development team of the BSES-SF and the team of the Erasmus Clinic Rotterdam which had previously translated the instrument into Dutch, the phrase “*Ich bin zuversichtlich, dass*” (which equates to “I am confident that”) was added at the beginning of every item in a deviation from the original text.

Cognitive debriefing

During the cognitive debriefing of the 15 women, a number of terms were identified which had to be reformulated due to comprehension problems arising from the translation.

Incontinence

The question which screens for incontinence also asks about involuntary flatulence. During the interviews it was found that the

majority of respondents were not familiar with the word *Flatulenz*. The phrase “*Abgang von Winden*” (= passing wind), which is used in the Jorge-Wexner Incontinence Score, was also not understood. After the term was paraphrased and respondents were asked to say what it meant in their own words, test persons repeatedly used the word “*Blähung*”. After further discussions with the participants in the study, the term “*Abgang von Luft*” (= passing gas) was identified as medically correct as well as generally intelligible.

Pain during sexual intercourse

In response to the question whether pain had an impact on the respondent’s sexual life, the participants in the study noted that the range of possible responses for persons who had no sexual activity was insufficient (“0 = In the past 30 days I did not experience pain”, “1 = Not at all” to “5 = Very much”). As the question was a validated item from the PROMIS Sexual Function item bank, a new item (“In the past 30 days, how satisfied were you with your sex life?”) was developed, based on item SFSAT201 of PROMIS, and prepended to the original item, with possible responses ranging from “0 = In the last 30 days I did not have an active sex life” to “1 = Not at all” and “5 = Very much”. The question about any impairments to the respondent’s sex life due to pain was subsequently only asked if the respondent had previously stated that she had an active sex life.

An overview of all amendments undertaken following the cognitive debriefing is given in ► **Table 2**.

Discussion

The ICHOM Set for Pregnancy and Childbirth covers a range of dimensions which are relevant for women during pregnancy and birth [9]. Because the three instruments BSES-SF, MIBS and BSS-R had not yet been translated into German, full use of the ICHOM Set was not possible in German speaking countries. As part of this study these three instruments have been translated into German, the comprehensibility of the translations was verified using a cohort of 15 pregnant women and women who had just given birth. Some of the wording was amended following their input.

The FACIT method was used for the translation process to ensure that the translation would be valid. The FACIT method is a qualitative method which is used to ascertain that the meaning of a translation corresponds to the meaning of the original text. It does not provide any information about the measurement capabilities of translated instruments, which is why an additional quantitative examination based on this study will be necessary.

Part of the qualitative validation consisted of the cognitive debriefing of 15 pregnant women and women who had just given birth, who were surveyed using the new translation. The women were initially sceptical about such a survey, which is regularly planned as part of clinical routine, but this turned into great interest and broad openness in the course of the survey. The women felt that the issues covered during the cognitive debriefing, such as expectations about breastfeeding, mother-child bonding and birth experience, were relevant. To successfully implement this instrument in future, it appears to be important to counter initially

► **Table 2** List of adaptations made to the translation after the validation process with comments.

Item and instrument in the original	Formulation of the question in German after the first translation	Changes made following debriefing	Comments
I can always breastfeed my baby without using formula as a supplement. (BSES03)	<i>Ich bin zuversichtlich, mein Kind immer stillen zu können und keine Ergänzung mit Milchpulver/ Formulanahrung (industriell hergestellte Anfangsnahrung) zu benötigen.</i>	<i>Ich bin zuversichtlich, mein Kind immer stillen zu können und keine Ergänzung mit industriell hergestellter Anfangsnahrung (Milchpulver) zu benötigen.</i>	The term <i>Formulanahrung</i> (= formula food) was not clear. <i>Industriell hergestellter Anfangsnahrung in Kombination mit Milchpulver</i> (= industrially produced infant formula) was understood.
I can always keep wanting to breastfeed. (BSES07)	<i>Ich bin zuversichtlich, dass ich auch weiterhin stillen möchte.</i> and <i>Ich bin zuversichtlich, so lange wie möglich stillen zu wollen.</i>	<i>Ich bin zuversichtlich, so lange wie möglich stillen zu wollen.</i>	A decision/vote was taken during cognitive debriefing about which wording was understood better.
I can always deal with the fact that breastfeeding can be time-consuming. (BSES10)	<i>Ich bin zuversichtlich, dass ich immer damit umgehen kann, dass das Stillen sehr zeitaufwendig ist.</i> and <i>Ich bin zuversichtlich, dass ich immer mit der Tatsache umgehen kann, dass das Stillen sehr zeitaufwendig ist.</i>	<i>Ich bin zuversichtlich, dass ich immer damit umgehen kann, dass das Stillen sehr zeitaufwendig ist.</i>	A decision/vote was taken during cognitive debriefing. The second sentence generated negative emotions in the study participants, the thought of work.
I can always continue to breastfeed my baby for every feeding. (BSES12)	<i>Ich bin zuversichtlich, mein Kind stets bei jeder Mahlzeit zu stillen.</i>	<i>Ich bin zuversichtlich, mein Kind stets bei jeder Mahlzeit (bei jeder Nahrungsaufnahme) stillen zu können.</i>	Question was amended as the term <i>Mahlzeit</i> was understood to refer to breakfast, lunch and dinner.
Obstetric history Bleeding so much during pregnancy, birth, or after giving birth that you needed to be given blood.	<i>Blutungen während der Schwangerschaft, Geburt oder nach der Geburt, die so stark waren, dass Sie Fremdblut bekommen mussten.</i>	<i>Blutungen während der Schwangerschaft, Geburt oder nach der Geburt, die so stark waren, dass Sie eine Bluttransfusion bekommen mussten.</i>	All of the women were much more familiar with the term <i>Bluttransfusion</i> (= blood transfusion) rather than <i>Fremdblut</i> (= allogeneic blood).
In the past month, have you leaked urine, leaked stool or passed gas by accident?	<i>Hatten Sie im letzten Monat ungewollt Verlust von Urin oder Abgang von Stuhl oder Winden?</i>	<i>Hatten Sie im letzten Monat ungewollt Verlust von Urin oder Abgang von Luft oder Stuhl?</i>	The term <i>Winden</i> was often not understood. Alternative suggestions included: <i>Abgang von Blähungen, unkontrollierte Blähung</i> ; most patients were unfamiliar with the medical term <i>Flatulenzen</i> .
In the past 30 days, how much has pain affected your satisfaction with your sex life?	<i>In den letzten 30 Tagen, wie stark hat Schmerz Ihre Zufriedenheit mit Ihrem Sexualleben beeinflusst?</i>	<i>In den letzten 30 Tagen, wie stark haben Schmerz in irgendeinem Teil Ihres Körpers Ihre Zufriedenheit mit Ihrem Sexualleben beeinflusst?</i>	For many affected women, it was not clear which pain was being referred to. The sentence was therefore amended to include <i>"in irgendeinem Teil Ihres Körpers"</i> (= in any part of your body).

expected skepticism through targeted communication about the aims and benefits of the survey.

The use of PROs in obstetrics should facilitate more personal doctor–patient communication which takes account of women's medical issues and insecurities. They also provide a quality assessment of the care received, which may lead to a new definition of the need for treatment once the patient's perspective has also been taken into account. The Birth Satisfaction Scale – Revised (BSS-R) is used to measure the satisfaction of a woman with the birth experience and her perception of the birth experience [19]. Contrary to longstanding assumptions, birth experience did not correlate with the pain experienced during the birth. Factors such

as the extent of support provided by medical staff and the involvement of the women giving birth in decision-making during the course of the birth appear to be significantly more relevant [20]. A woman's birth experience has a significant impact on her subjective perception of health and health-related quality of life [21] and thus also has an impact on the development of the mother and child relationship [22]. A healthy and secure bond between mother and child is important for the psychological and physiological development of the child [23]. The MIBS focuses particularly on the early emotional bond between mother and newborn; because of its short structure and its focus on just eight emotion-laden adjectives, it can also be administered to women with lower

language skills [24, 25]. Early identification of an attachment disorder between mother and child can reduce the risk of physical and mental abuse [26].

The translation of the Breastfeeding Self-Efficacy Scale – Short Form (BSES-SF) also offers the opportunity to improve care as it allows risk factors which hamper breastfeeding to be identified early on. According to the recommendations of the World Health Organization (WHO), infants should be exclusively breastfed for the first six months of life and breastfeeding on demand should continue at least until the age of two years [27]. With an initial breastfeeding rate of 82% just after birth, Germany is at the lower end of the scale in international comparisons [28]. Data from the 2018 Nutrition Report by the German Nutrition Society showed that 56% of infants in Germany are exclusively breastfed up to their 4th month of life and that this figure drops to just 8% by their 6th month of life [29]. There are many risk factors which may lead to women not breastfeeding their infants. In addition to socio-demographic factors, risk factors include lack of the intention to breastfeed or a limited wish to breastfeed, lack of a positive attitude to breastfeeding as well as the giving of supplementary liquids to the neonate in the first 2 weeks of life. These specific factors are measured by the BSES-SF, and the information obtained provides medical staff with the opportunity to offer targeted breastfeeding advice to pregnant women at an early stage.

As the instruments BSES-SF, MIBS and BSS-R have been translated into German, German hospitals now have the option to implement the complete ICHOM Set for Pregnancy and Childbirth and compare PRO results with those of other hospitals worldwide. To do this, it will be necessary to adapt the time points at which the surveys are administered. In Germany, the surveys at the time points “early 3rd trimester” and “within 3 days postpartum” can usually be carried out by the institution where the woman gives birth, the surveys at the time points “1st postpartum checkup” and “6 months postpartum” can be initiated and evaluated digitally by the hospital where the woman gives birth. To be able to administer the initial survey in the 1st trimester of pregnancy, a cross-sectoral survey, which would require the participation of non-hospital-based doctors, is conceivable in future and would be desirable. The ICHOM Standard Set for Pregnancy and Childbirth provides the opportunity to obtain a cross-sectional view of the health and quality of life of women, from their first trimester of pregnancy up until 6 months postpartum, and to record the quality of care across a number of care sectors.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

- [1] WHO. WHO recommendations: Intrapartum care for a positive childbirth experience. Geneva: World Health Organization; 2018:200
- [2] Deutsche Gesellschaft für Gynäkologie und Geburtshilfe e.V. Leitlinienprogramm der DGGG e.V., OEGGG e.V. und SGGG e.V. Accessed May 01, 2020 at: www.dggg.de/leitlinien-stellungnahmen/leitlinien/
- [3] Porter ME, Teisberg EO. Redefining Health Care: Creating Value-based Competition on Results. Boston, Mass.: Harvard Business School Press; 2006
- [4] NHS; Health and Social Information Centre. Secondary Care Analysis (PROMs) NHS Digital. Patient Reported Outcome Measures (PROMs) in England – A guide to PROMs methodology. 2017. Accessed May 01, 2021 at: <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/patient-reported-outcome-measures-proms>
- [5] Ministerie van Volksgezondheid Welzijn en Sport. Outcome based healthcare 2018–2022. Den Haag: Government of the Netherlands; 2018
- [6] U.S. Food and Drug Administration. 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. doi:10.1186/1477-7525-4-79
- [7] Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. JAMA 1995; 273: 59–65
- [8] 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: 1179–1194. doi:10.1016/j.jclinepi.2010.04.011
- [9] Nijagal MA, Wissig S, Stowell C et al. Standardized outcome measures for pregnancy and childbirth, an ICHOM proposal. BMC Health Serv Res 2018; 18: 953–953. doi:10.1186/s12913-018-3732-3
- [10] Jorge JM, Wexner SD. Etiology and management of fecal incontinence. Dis Colon Rectum 1993; 36: 77–97. doi:10.1007/bf02050307
- [11] Banerjee C, Banerjee M, Hatzmann W et al. The German version of the ‘ICIQ Vaginal Symptoms Questionnaire’ (German ICIQ-VS): an instrument validation study. Urol Int 2010; 85: 70–79. doi:10.1159/000316337
- [12] Hays RD, Björner JB, Revicki DA et al. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. Qual Life Res 2009; 18: 873–880. doi:10.1007/s11136-009-9496-9
- [13] Löwe B, Spitzer RL, Zipfel S et al. Gesundheitsfragebogen für Patienten (PHQ-D). Manual und Testunterlagen. 2. Aufl. Karlsruhe: Pfizer; 2002
- [14] Bergant AM, Nguyen T, Heim K et al. Deutschsprachige Fassung und Validierung der “Edinburgh postpartum depression scale” [German validation of the Edinburgh Postpartum Depression Scale]. Dtsch Med Wochenschr 1998; 123: 35–40. doi:10.1055/s-2007-1023895
- [15] Eremenco SL, Cella D, Arnold BJ. A comprehensive method for the translation and cross-cultural validation of health status questionnaires. Eval Health Prof 2005; 28: 212–232. doi:10.1177/0163278705275342
- [16] Willis G. Analysis of the Cognitive Interview in Questionnaire Design. New York: Oxford University Press; 2015
- [17] Konrad K. Laut Denken. In: Mey G, Mruck K, Hrsg. Handbuch Qualitative Forschung in der Psychologie. Wiesbaden: Springer; 2010: 476–490
- [18] Nespoli A, Colciago E, Pedroni S et al. The Birth Satisfaction Scale-Revised (BSS-R): process of translation and adaptation in an Italian context. Ann Ist Super Sanita 2018; 54: 340–347. doi:10.4415/ann_18_04_11
- [19] Hollins-Martin CJ, Martin C. Development and psychometric properties of the Birth Satisfaction Scale-Revised (BSS-R). Midwifery 2014; 30: 610–619
- [20] Hodnett ED. Pain and women’s satisfaction with the experience of childbirth: A systematic review. Am J Obstet Gynecol 2002; 186: S160–S172. doi:10.1016/S0002-9378(02)70189-0
- [21] Schäfers R. Einfluss des Geburtserlebens auf die subjektive Gesundheitseinschätzung von Frauen. GMS Z Hebammenwiss 2015; 2: Doc02. doi:10.3205/zhw000005
- [22] Rossen L, Mattick RP, Wilson J et al. Mother-Infant Bonding and Emotional Availability at 12-Months of Age: The Role of Early Postnatal Bonding, Maternal Substance Use and Mental Health. Matern Child Health J 2019; 23: 1686–1698. doi:10.1007/s10995-019-02809-1

- [23] Klaus MH, Kennell JH. Parent/Infant Bonding by Marshall H. Klaus. Michigan: Mosby; 1982
- [24] Taylor A, Atkins R, Kumar R et al. A new Mother-to-Infant Bonding Scale: links with early maternal mood. *Arch Womens Ment Health* 2005; 8: 45–51. doi:10.1007/s00737-005-0074-z
- [25] van Bussel JC, Spitz B, Demyttenaere K. Three self-report questionnaires of the early mother-to-infant bond: reliability and validity of the Dutch version of the MPAS, PBQ and MIBS. *Arch Womens Ment Health* 2010; 13: 373–384. doi:10.1007/s00737-009-0140-z
- [26] Brockington JF, Aucamp HM, Fraser C. Severe disorders of the mother–infant relationship: definitions and frequency. *Arch Womens Ment Health* 2006; 9: 243–251. doi:10.1007/s00737-006-0133-0
- [27] Michaelsen KF, Weaver L, Branca F et al. Stillen und Ernährung von Neugeborenen und Kleinkindern. WHO 2003; 87: 146. Accessed November 17, 2021 at: https://www.euro.who.int/__data/assets/pdf_file/0004/98302/WS_115_2000FE.pdf
- [28] von der Lippe E, Brettschneider AK, Gutsche J et al. [Factors influencing the prevalence and duration of breastfeeding in Germany: results of the KiGGS study: first follow up (KiGGS Wave 1)]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2014; 57: 849–859. doi:10.1007/s00103-014-1985-5
- [29] Kersting M, Hockamp N, Burak C et al. Studie zur Erhebung von Daten zum Stillen und zur Säuglingsernährung in Deutschland – SuSe II. In: *Deutsche Gesellschaft für Ernährung, Hrsg. 14. DGE-Ernährungsbericht. Vorveröffentlichung Kapitel 3. Bonn (2020) V1–V34*

ERRATUM**Das ICHOM-Standardset für Schwangerschaft und Geburt – Übersetzung und sprachliche Anpassung für Deutschland**

Claudia Hartmann, Matthias Rose, Alexander Weichert et al.
Geburtsh Frauenheilk 2022; 82: 747–754
 10.1055/a-1666-0429
 online publiziert: 07.07.2022

Im oben genannten Artikel ist im Literaturverzeichnis die Literaturstelle 19 nicht richtig angegeben. Korrekt ist: Hollins-Martin CJ, Martin C. Development and psychometric properties of the Birth Satisfaction Scale-Revised (BSS-R). *Midwifery* 2014; 30: 610–619. Die Korrektur wurde in der Onlineversion des Artikels ausgeführt am: 22.07.2022



PROMIS-29 and EORTC QLQ-C30: an empirical investigation towards a common conception of health

Claudia Hartmann¹ · Felix Fischer¹ · Christoph P. Klapproth¹ · Robert Röhle² · Matthias Rose^{1,3} · Maria M. Karsten²

Accepted: 7 December 2022 / Published online: 9 January 2023
 © The Author(s) 2023

Abstract

Purpose The assessment of health-related quality of life (HRQOL) measured via patient-reported outcomes (PROs) is a key component in clinical trials and increasingly used in clinical routine worldwide. Two PRO measures (PROMs) that share the same definition of health and report outcomes on a comparable T-metric anchored to general population samples are the PROMIS-29 and the EORTC QLQ-C30. In this study, we investigate the empirical agreement of these underlying concepts.

Methods We collected PROMIS-29 and EORTC QLQ-C30 data from 1,478 female patients at a breast cancer outpatient centre. We calculated descriptive statistics and correlations between the subscales of both instruments. We performed exploratory (EFA) and confirmatory factor analysis (CFA) in randomly split subsamples in order to assess the underlying psychometric structure of both instruments.

Results The cohort (mean age = 47.4, ± 14.49) reported comparable mean HRQOL scores between the corresponding subscales of both instruments similar to general population reference values. Correlation between the corresponding subscales of both instruments ranged between 0.59 (Social Role) and 0.78 (Physical Functioning). Both an exploratory and a theoretically driven confirmatory factor analysis provided further support for conceptual agreement of the scales.

Conclusion EORTC QLQ-C30 and PROMIS-29 showed similar scores and satisfactory agreement in conceptual and statistical analysis. This suggests that the underlying conceptualization of health is reasonably close. Hence, the development of score transformation algorithms or calibration of both instruments on common scales could prospectively increase the comparability of clinical and research PRO data collected with either instrument.

Keywords PROMIS-29 · EORTC QLQ-C30 · Health-related quality of life · Conceptualization of health · Comparable outcome

Purpose

Collecting patient-reported outcomes (PROs) to measure health-related quality of life (HRQOL) in clinical routine has become more common and accepted worldwide, as recent studies have shown that a systematic collection of PROs as

a monitoring tool can improve overall survival and reduce hospitalization and emergency visits [1]. Furthermore, routine collection of PRO data improves patient–physician communication and can lead to better outcomes [2–4].

PROs collected in clinical routine allow clinicians to compare outcomes of similar patient groups between countries and to improve overall health care quality [5]. Currently, over 3,900 PRO instruments have been developed and are available to measure HRQOL [6]; the greatest challenge for the upcoming years is to standardize PRO assessments with common underlying latent constructs such as depression, physical function, fatigue, etc.

Two PRO instruments that are widely used to assess HRQOL in a comprehensive manner are the Quality of Life Questionnaire-Core30 (QLQ-C30), developed by the European Organization for Research and Treatment Cancer

✉ Claudia Hartmann
claudia.hartmann@charite.de

¹ Department of Psychosomatic Medicine, Center of Internal Medicine and Dermatology, Charité - Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany

² Department of Gynecology With Breast Center, Charité University Hospital Berlin, Berlin, Germany

³ Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, USA

(EORTC), and the PROMIS-29 as part of the Patient-Reported Outcome Measurement Information System (PROMIS). The QLQ-C30 was developed as a core instrument for patients with any cancer. Therefore, the instrument contains relevant items covering generic general health as well as items covering common symptoms following cancer therapy (e.g. nausea, appetite loss) [7]. The PROMIS-29 was developed to capture PROs for a wide range of chronic illnesses and disorders [8], and therefore covers only domains of general health (e.g. physical function, pain). Both instruments share the same definition of health as a multidimensional construct of dimensions such as physical function, mental health, social participation, sleep, and pain [7, 9]. Both are also based on the methodical and conceptual framework by Ware et al. [10, 11] to assess the health status of patients. A main difference between those two measures is the population they were developed for. Nevertheless, the PROMIS-29 has been successfully used in cancer populations [12–14] and is compared to the QLQ-C30 equivalent in terms of value and usefulness for patients and clinicians [15].

Both instruments are centrepieces of a wider framework of health assessment and have each been translated and validated in more than 40 languages. While PROMIS measures were developed using item response theory (IRT) methods, EORTC initially used methods of classical test theory for questionnaire development and recently adopted IRT to develop item banks for all QLQ-C30 domains. IRT allows a more flexible and efficient measurement of the targeted domains, e.g. using computer adaptive tests (CAT) or creating specific shortforms [16]. IRT is also used to score each domain measured on a standardized T-score metric, where 50 describes the mean (M) and 10 the standard deviation (SD). For both instruments, 50 represents the mean score in the general population in the US and Europe showing no significant differences between the population. [17] [18] [19] Using such a T-score metric makes similar domains from different instruments directly comparable [20].

One example that highlights the similarities and shared conception underlying both instruments is that the U.S. Food & Drug Administration (FDA) recommends both the EORTC and the PROMIS Physical Function scales in its Guidance Document “Core Patient-Reported Outcomes in Cancer Clinical Trials” [21]. As a first step to allow direct comparability of the results measured with the QLQ-C30 and PROMIS-29, the aim of this paper is to (1) descriptively assess the similarity of the PROMIS-29 and the EORTC QLQ-C30 at the domain and item levels, (2) empirically assess the agreement of comparable domain scores in terms of mean scores and correlations, and (3) assess whether a common underlying factor structure for both measures can be established. We expect that the mean scores of similar domains are reasonably close and that the correlation

coefficients show a strong association (>0.6) between convergent domains. The following exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) will further investigate the underlying psychometric structure of both instruments.

Methods

Sample

PROMIS-29 and QLQ-C30 data were collected from patients visiting the breast cancer outpatient clinic at Charité – Universitätsmedizin Berlin during routine diagnosis between November 2016 and March 2021. We included the first PRO assessment that was taken at the initial visit from 1,478 female ambulant patients for our analysis regardless of the breast disease (breast cancer, fibroadenoma and other). To avoid missing responses, the digital survey did not allow skipping questions.

Measures

The QLQ-C30 is a cancer-specific and core instrument in the library of EORTC instruments focussing on the general HRQOL and was first mentioned in 1986 [7, 22]. It comprises five functional scales, social (SF), emotional (EF), cognitive (CF), role (RF) and physical (PF), plus 8 symptom scales, pain (PA), fatigue (FA), dyspnoea (DY), sleep disturbance (SD), appetite loss (AL), constipation (CO), diarrhoea (DI) and nausea & vomiting (NV). Additionally, the financial status (FD), a global quality of life and global health scale (QOL and GLQ) as well as several symptom measures are collected. In total, the QLQ-C30 uses 30 items. Responses to each item are made on a four-point Likert scale (1 = “Not at all”, 2 = “A little”, 3 = “Quite a bit”, 4 = “Very much”) for Items 1–28 and a seven-point Likert scale (1 = “Very poor” to 7 = “Excellent”) for Items 28 and 29 (global health and quality of life) [7]. The recall period is not specified for the first 5 items, and all the following 25 items use a week as the recall period.

The PROMIS was funded in 2004 by the United States under the National Institute of Health (NIH) Roadmap for Medical Research Initiative [9] and since then, developed over 100 item banks covering different aspects of physical, mental and social health. The PROMIS-29 Profile—a 29-item instrument—combines short assessments of eight core constructs of HRQOL: physical function (PF), sleep disturbance (SD), pain interference (PI) and pain intensity (PIN), fatigue (FA), anxiety (AN), depression (DE) and ability to participate in social roles and activities (SRAA). The instrument includes 8 subscales and a five-point Likert scale where 3 scales (AN, DE, SRAA) use 5 = “Never”

to 1 = “Always” and 3 scales (PI, SD, FA) use 5 = “Not at all” to 1 = “Very much” [19]. The PF Likert scale ranges from 5 = “Without any difficulty” to 1 = “unable to do”. Pain intensity is measured on a 10-Point visual analogue scale (VAS) from 0 = “No pain” to 10 = “Worst pain imaginable”. All domains use a 7-day recall period, except for PF and SRAA. [8, 19]. Table 1 shows the scale and item content for both instruments.

Analysis

We summarized the demographic characteristics of the sample. We then transformed the individual item responses for each domain to *T*-scores for both the QLQ-C30 and the

PROMIS-29 using algorithms provided by the EORTC QLQ group and the German PROMIS National Center. Means and standard deviations for the PROMIS-29 and QLQ-C30 domain scores were computed. We compared these scores with general population reference data reported in the previous papers [17, 18]. We assessed correlations between all domains using the Pearson correlation coefficient. We defined a correlation > 0.6 as high and > 0.4 as moderate. [23] Furthermore, we performed an exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) to evaluate the dimensional structure of both instruments.

The combined dataset was randomly split in half, resulting in an equal amount of $N = 739$ for EFA and CFA. We investigated the number of factors to extract in the EFA

Table 1 Domains and items of the PROMIS-29 and the EORTC QLQ-C30

PROMIS-29		QLQ-C30	
Domain	Item	Domain	Item
Physical function	Ability to do chores such as vacuuming or yard work	Physical function	Strenuous activities, like carrying a heavy shopping bag or a suitcase
	Ability to run errands and shop		Take a long walk
	Ability to go for a walk of at least 15 min		A short walk outside of the house
	Ability to go up and down stairs at a normal pace		Stay in bed or a chair during the day
Anxiety	Feel fearful	Emotional function	Help with eating, dressing, washing yourself or using the toilet
	Hard to focus on anything other than anxiety		Worry
	Overwhelmed by worries		Feel tense
	Feel uneasy		Feel depressed
Depression	Feel worthless	Emotional function	Feel irritable
	Feel helpless		
	Feel depressed		
	Feel hopeless		
Pain	Pain interferes with day-to-day activities	Pain	Had pain
	Pain interferes with work around the home		Pain interference with daily activities
	Pain interferes with ability to participate in social activities		
	Pain interferes with household chores		
Fatigue	Pain scale	Fatigue	Need to rest
	Feel fatigued		Feel weak
	Trouble starting things because I am tired		Tired
	Average feeling of run-down		
Sleep	Average fatigue	Sleep	Trouble sleeping
	Sleep quality		
	Sleep was refreshing		
	Problem with sleep		
Social activity	Difficulty falling asleep	Role function	Doing either work or other daily activities
	Regular leisure activities with others		Pursuing hobbies or other leisure time activities
	Family activities that I want to do	Social function	Physical condition or medical treatment interfered with family life
	Usual work (including work at home)		Physical condition or medical treatment interfered with social activities
	Activities with friends that I want to do		

dataset by assessing eigenvalues using the Kaiser criterion. Since this tends to overestimate the number of factors to extract, we also performed a scree test and a parallel analysis. We performed ordinary least squares factor analysis based on the polychoric correlation matrix to account for the ordinal nature of the items and used oblimin rotation to account for the correlated nature of the constructs. We then compared multiple factor solutions exploratively by assessing the resulting patterns of factor loadings.

For the CFA, we estimated an a priori specified model based on our content analysis and the results from the EFA. For this model, we assigned items of the same construct from both the PROMIS-29 and QLQ-C30 to the correlated latent factors. To account for the ordinal nature of the responses, we used the WLSMV estimator. We report standardized model parameter estimates and assessed the comparative fit index (CFI), the root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR) to evaluate model fit (criteria RMSEA < 0.08, CFI > 0.95, SRMR < 0.08 [24]). Modification indices were descriptively assessed to investigate potential misspecifications of the model.

Symptom measures of the QLQ-C30 as well as the FD, CF, NV and global scales were excluded a priori from the correlation analysis, EFA and CFA. FD was excluded as it is neither a measure of function nor a symptom. The symptoms and global scales were excluded since they are not covered by the PROMIS-29. We excluded the PROMIS Pain intensity item, as it is a numerical analogue scale not used in latent variable modelling.

All analyses were conducted with R version 4.0.2 software using the following packages: lavaan (0.6–8), nFactors (2.4.1), tidyr (1.1.2) and dplyr (1.0.3).

Results

Content comparison

The content assessment of both instruments showed a similar conceptualization of health on the domain as well as on the item level (Table 1). Each of the seven domains of the PROMIS-29 is assignable to one domain of the QLQ-C30. The PROMIS-29 domains AN and DE are linked to the EF domain of QLQ-30, and the SRAA domain combines the two QLQ-C30 domains RF and SF. A main difference between both instruments is the number of items used to measure a domain. Whereas the PROMIS-29 uses 4 items for each domain, the QLQ-C30 uses a different number of items to measure a single construct: for example, one item for SF, 3 items for FA and 5 items for PF (see Table 1).

In addition to the comparability of the content, the wording used in individual items further shows similarities

between the two instruments. The items measuring depression both use the wording “feel depressed”/“felt depressed”, and for anxiety, both use “worry”/“worries”. Similar wording can also be seen for items in the domains FA, PI and SD. Above all, there are differences that are particularly evident within the spectrum of the domains. Looking at the domain PF, the QLQ-C30 items cover a wider range from very basic (eating, dressing), easy (short walk) and difficult tasks (long walk), whereas the spectrum of the PROMIS-29 does not cover basic tasks. The PROMIS-29 domain SL, however, covers a broader spectrum of sleep quality than the single item used in the QLQ-C30.

Looking at the overall wording items of the generic instrument, PROMIS-29 does not reflect any specific disease. Despite QLQ-C30 being developed as a cancer-specific instrument, no reference to a specific disease is made in item wording.

Descriptive statistics

The mean age of the 1,478 female patients in the sample was 47.4 years (SD = 14.49, range 15–90), with 41.54% of the women being diagnosed with malignant and 26.86% with benign breast cancer. Most patients (65.97%) had a high school diploma or further education. Demographics are summarized in Table 2.

Table 2 Demographic characteristics of the sample

Patient characteristics (N = 1,478)		
Characteristics	No	%
<i>Age at enrolment</i>		
Mean		47.4
Range		15–90
<i>Age group</i>		
<30	192	12.99%
30–40	312	21.11%
41–50	391	26.45%
51–60	309	20.91%
61–70	177	11.98%
>70	97	6.56%
<i>Education group</i>		
High	975	65.97%
Middle	369	24.97%
Low	128	8.66%
not specified	6	0.41%
<i>Diagnose group</i>		
Mamma-CA	572	38.70%
Fibroadenoma	225	15.22%
Benign	397	26.86%
DCIS	42	2.84%
Second opinion	242	16.37%

Table 3 shows the *T*-scores of both instruments by domain. We observed no relevant deviation regarding HRQOL of our cohort from the general population, as the mean scores of all domains were close to 50 ($M = 50$, $SD \pm 10$) [17, 18]. An exception is mental health, where we observed elevated levels of anxiety (PROMIS) and reduced levels of emotional functioning (EORTC). Overall, the mean *T*-scores of the corresponding domains of both measures show only small differences (< 5) between each other. More descriptives results (mean, standard deviation, normdata, kurtosis, skewness) of the included health domains can be found in the electronic supplement Table I.

Correlation analysis

All PROMIS-29 domains correlate highly (> 0.6) with at least one other domain of the QLQ-C30. Most of the similar descriptive domains show the highest correlation to each other except for the QLQ-C30 domain RF, with a correlation to the corresponding domain of 0.59 and a higher correlation to PF and PI of the PROMIS-29 (0.68 and -0.63). As expected from the content analysis, the QLQ-C30 domain EF shows high correlations with the PROMIS-29 domains AN (-0.69) and DP (-0.70) and with FA (-0.62). All coefficients are summarized in Table 4.

The multitrait-multimethod analysis showed a high average correlation of 0.73 for conceptually similar domains. The average correlation attributed to the method is similar, looking only at PRMIS-29 and QLQ-30, with 0.52 and 0.51, respectively. The average correlation for conceptually

distinct domains that do not share the same method is lowest at 0.47.

Exploratory factor analysis (EFA)

The EFA resulted in 6 eigenvalues greater than 1 (19.95, 5.08, 3.02, 1.85, 1.39, 1.14), indicating 6 factors to extract according to the Kaiser criterion. Parallel analysis also suggested 6 factors. An inspection of the scree plot suggests 3–5 factors. We, therefore, estimated four different EFA models with 3 to 6 factors.

We found the EFA with 5 factors to be interpretable as the most meaningful. These 5 factors were categorized as Physical Function, Mental Health, Fatigue, Pain and Sleep. These factors explain 72% of the total variance. All items of both instruments related to pain and sleep load on the same factor. All PROMIS-29 domains and associated items show the highest loading to one modelled factor with loadings of 0.6/–0.6 and higher/lower, respectively, except all items of the SRAA domain, which have loadings between -0.56 and -0.51 to the factor Physical Function. Eight out of 19 items of the QLQ-C30 did show a relevant loading for a single factor, and six showed a moderate or single relevant loading assignable to one factor. Items of the QLQ-C30 domain EF and SF have cross-loadings to several factors, whereas two QLQ-C30 EF items show the highest loadings on the factor Fatigue (0.51, 0.49) and two on the factor Mental Health (0.55, 0.51). All factor loadings of the Exploratory Factor Analysis with 5 factors can be found in the electronic supplement Table II.

Table 3 Results of PROMIS-29 and QLQ-C30 compared to norm data

		PROMIS-29		QLQ-C30	
		<i>T</i> -Score Mean (SD)	Norm	<i>T</i> -Score Mean (SD)	Norm
Mental health	Anxiety	56.19 (9.82) ^d	53.1		
	Depression	51.52 (8.94) ^d	52.7		
	Emotional function			44.17 (9.72) ^a	50.63
	Fatigue	47.85 (10.11) ^d	48.3	51.52 (10.15) ^b	51.17
Physical health	Physical function	51.17 (7.61) ^c	51.4	50.68 (9.74) ^a	48.33
	Pain Interference	48.79 (8.93) ^d	52.4	50.18 (9.03) ^b	50.84
	Sleep disturbance	50.43 (8.95) ^d	49.2	54.22 (8.41) ^b	50.96
Social health	Social roles and activities	54.04 (9.63) ^c	50.8		
	Role function			48.96 (9.04) ^a	49.13
	Social function			47.86 (8.76) ^a	49.93

^aQLQ-C30 domains in *T*-score, 50 = population average, $-/+ 10$ SD worse/better than population average, i.e. higher values indicate better function

^bQLQ-C30 domains in *T*-score, 50 = population average, $-/+ 10$ SD better/worse than population average, i.e. lower values indicate better function

^cPROMIS domains in *T*-score, 50 = population average, $-/+ 10$ SD worse/better than population average, higher values indicate better function

^dPROMIS domains in theta *T*-score, 50 = population average, $-/+ 10$ SD better/worse than population average, lower values indicate better function

Table 4 Correlation coefficients of the PROMIS-29 and QLQ-C30 domains, highlighted correlations >0.6 / <-0.6 and multitrait-multimethod correlation matrix.

		PROMIS-29							QLQ-C30					
		Anxiety	Depression	Fatigue	Pain Interference	Physical Function	Sleep Disturbance	Social roles & activities	Physical Function	Emotional Function	Role Function	Social Function	Fatigue	Pain
PROMIS-29	Depression		0.76											
	Fatigue		0.49	0.64										
	Pain Interference		0.32	0.43	0.54									
	Physical Function		-0.24	-0.37	-0.49	-0.73								
	Sleep Disturbance		0.43	0.49	0.47	0.37	-0.32							
	Social roles and activities		-0.48	-0.62	-0.70	-0.63	0.62	-0.42						
QLQ-C30	Physical Function		-0.19	-0.33	-0.47	-0.64	0.78	-0.32	0.59					
	Emotional Function		-0.69	-0.70	-0.62	-0.39	0.34	-0.47	0.57	0.32				
	Role Function		-0.24	-0.35	-0.49	-0.63	0.68	-0.29	0.59	0.62	0.37			
	Social Function		-0.48	-0.56	-0.55	-0.53	0.54	-0.37	0.66	0.48	0.61	0.57		
	Fatigue		0.35	0.49	0.75	0.58	-0.59	0.43	-0.65	-0.59	-0.57	-0.59	-0.56	
	Pain		0.26	0.35	0.47	0.75	-0.63	0.35	-0.51	-0.61	-0.38	-0.60	-0.47	0.60
	Sleep		0.39	0.45	0.46	0.40	-0.38	0.77	-0.44	-0.37	-0.51	-0.38	-0.42	0.54

average correlations

same trait - different method 0.73

same method - different trait - PROMIS-29 0.52

same method - different trait - QLQ-C30 0.51

different method - different trait 0.47

Confirmatory factor analysis (CFA)

Content analysis showed 7 domains with corresponding items from both measures. Based on the EFA, we modelled anxiety and depression as common factors. Figure 1 shows the resulting CFA model, which had a satisfactory fit of the data according to RMSEA = 0.074, CFI = 0.968 and SRMR = 0.074. Overall, we found that all items had high loadings (> 0.6) on the hypothesized factors and that correlation between factors was as expected. The top modification indices suggest additional common variance not explained by our hypothesized model between the QLQ-C30 question 10 (“Did you need to rest?”) and the physical function and pain factors. Additionally, the PROMIS-29 item “I found it hard to focus on anything other than my anxiety” showed additional relations to factors of fatigue, social role, physical function and pain. Since the model fit was satisfactory, we did not model these additional associations.

Discussion

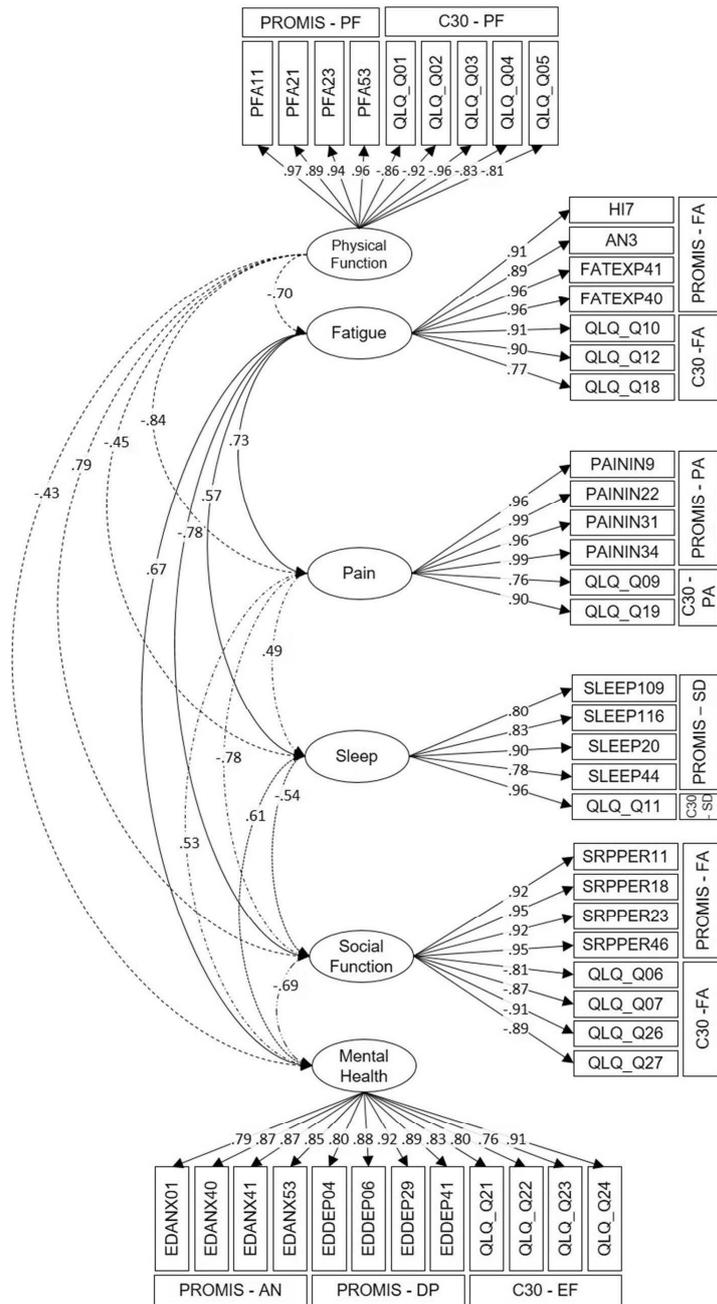
This paper focuses on the comparability of the two instruments QLQ-C30 and PROMIS-29 and their similar domains to prospectively increase the comparability of clinical and research PRO data collected with either instrument. In this study, we descriptively assessed the underlying HRQOL model of the PROMIS-29 and QLQ-C30 before we evaluated the similarity of both instruments empirically in a sample of 1,549 patients visiting the outpatient clinic for breast cancer and assessed whether a common underlying factor structure for both measures could be established. Our results

demonstrate that the underlying conceptualization of health is comparable in both instruments.

First, we found a descriptive similarity of both instruments on the domain as well as on the item level indicated by similar wording in domain titles and item descriptions. Despite the different patient population backgrounds of the two instruments—cancer patients in clinical trials vs. patients with chronic diseases—there are structural similarities in the choice of health domains, implying a similar understanding of health. This finding was expected because both are core instruments within their frameworks to measure general health.

Second, we demonstrated this similarity empirically by comparing the T-scores and evaluating the correlation coefficients of the subscales from both measures. It could be shown that subscales with similar descriptive content show similar T-scores and a stronger correlation than others. It can also be assessed that the domains PF and PI of the PROMIS-29 both show a strong correlation to the QLQ-C30 domains PF, RF and PA, indicating a strong association between those dimensions. The highest correlation was observed between the corresponding scales, but this finding shows that factors such as pain are interrelated or dependent with the outcome of factors such as PF and RF, which is particularly relevant for the interpretation of the results in practice. These findings are in line with the previous research [25, 26]. We also compared the T-scores to data from the general population and showed that they are similar, indicating that the results of our paper are not disease specific. This similarity between the outcomes of our sample and the general population is expected since data were collected prior to treatment and final diagnosis. Therefore, we would not

Fig. 1 Confirmatory factor analysis—loadings of all items and factors to one another



expect major differences in the general population. What can be expected is a slightly higher score in anxiety compared to norm data due to upcoming diagnosis and/or therapy decisions. This was measured well by the PROMIS-29. Differences between the two scores may be due to the different underlying norm samples (EU and US) and the measurement accuracy of the instruments and subscales. For example, SD is measured with a single item by the QLQ-C30 and four items by the PROMIS-29.

Based on the conceptual overlap between the two instruments, we empirically developed a 5 factor model in an EFA framework. Unlike previous papers, our focus was not on the validation of the instruments and their subscales in a certain population. We focussed on analysing whether these two measures are part of a greater model, hence using the same latent constructs to measure HRQOL. For the PROMIS-29 items, this model shows that 24 of 28 items can be assigned to one of the factors. For the QLQ-C30, only 8 out of 19 can be clearly assigned to one factor. An assignment of the other items or even the domain to one single construct was not always clear, which is also the case in a model with fewer factors. Similar results were found in earlier research [26]. Taking into account the appropriate fit of a theoretically derived model in the CFA, it can be assumed that the PROMIS-29 and QLQ-C30 share a similar conceptual model of health as perceived by patients. However, in both the EFA and CFA, some theoretically distinct constructs were inseparable or showed high correlations. For example, Social Function and Physical Function constituted a single factor in the EFA and were highly correlated ($r=0.79$) in the CFA. This suggests less than perfect discriminant validity.

Sharing the same metric is a huge advantage in comparing the outcome measured with two different instruments, the exchange of results, the clinical applicability and the reduction of the ambiguity of the interpretation [20]. Although both instruments share the same metric, the value 50 does not correspond to the 50 of the other instrument, since they have been calculated based on different patient populations (European for QLQ-C30 and US for PROMIS-29). However, as our and other research [18] showed, differences to the mean are minor, and comparison of scores is possible, with the general advantage outweighing the minor differences [20]. However, differences in the calibration samples could potentially mask real differences. Therefore, to allow precise comparisons between assessments made with both instruments, a common measurement model based on the same population is necessary.

Strengths and limitations

A major limitation of this study is that the sample is a convenience sample. This study evaluated data only from one sex and only from patients with breast-related diseases.

Therefore, we cannot generalize our findings to men or patients with other diseases. However, since all domains of both scores in our sample show no or only little deviation compared with the general population, it can be assumed that this results in a general validity.

A further limitation is derived from the number of items used to measure each construct. Since both measures are streamlined for rapid, individual, and comprehensive HRQOL, assessment results scoring leads to limited precision. To develop common measurement models across both instruments, it would be necessary to include enough items to adequately represent the full range of each construct with a reasonable measurement precision.

Finally, we did not develop conversion tables between PROMIS-29 and EORTC QLQ-C30 since both questionnaires only include very few items of each item bank, which are likely not reflecting the broad range of the underlying constructs and therefore score conversion could be biased.

Despite these limitations, our work has strengths, such as a large dataset with over 1,500 patients in a relevant population. Furthermore, the results from this sample have been compared to the general population, showing no major differences between both groups prior to treatment.

Conclusion

The EORTC QLQ-C30 and PROMIS-29 are two instruments that measure HRQL with a focus on different populations, which is why the QLQ-C30 covers a broader spectrum of dimensions and the instruments shall not be seen as redundant. Our research showed similar scores and satisfactory agreement in conceptual and statistical analysis between the two instruments. This suggests that the underlying conceptualization of health is reasonably close. Hence, the development of score transformation algorithms or calibration of overall HRQOL domains of both instruments on common scales could prospectively increase the comparability of clinical and research PRO data collected with either one and increase flexibility when choosing an instrument. The datasets analysed during the current study are not publicly available due to data privacy regulations but can be made available from the corresponding author on reasonable request.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11136-022-03324-7>.

Authors contributions CH: mainly wrote the script, created the figures and tables, compared both instruments at the domain and item levels and performed the empirical assessment. FF: designed the empirical assessment, wrote the code for the EFA and CFA and interpreted the results. CP,K: wrote the discussion and supported writing the script. RR: extracted the data from the outcome measurement platform and prepared the data for data analysis. MR: designed and drafted the script

idea as well as tables and figures of the work. MM.K: was in charge of data collection and revision of the script. All authors read and approved the final manuscript. All authors commented on the previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. None.

Data availability The datasets analysed during the current study are not publicly available due to data privacy regulations but can be made available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The Ethics Committee of the Campus Benjamin Franklin of the Charité Universitätsmedizin Berlin approved the study with the reference number EA4/127/16.

Consent for publication Not applicable.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Basch, E., Deal, A. M., Dueck, A. C., Scher, H. I., Kris, M. G., Hudis, C., & Schrag, D. (2017). Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. *JAMA*, *318*(2), 197–198.
- Velikova, G., Booth, L., Smith, A. B., Brown, P. M., Lynch, P., Brown, J. M., & Selby, P. J. (2004). Measuring quality of life in routine oncology practice improves communication and patient well-being: A randomized controlled trial. *Journal of Clinical Oncology*, *22*(4), 714–724.
- Detmar, S. B., Müller, M. J., Schormagel, J. H., Wever, L. D., & Aaronson, N. K. (2002). Health-related quality-of-life assessments and patient-physician communication: A randomized controlled trial. *JAMA*, *288*(23), 3027–3034.
- Snyder, C. F., & Aaronson, N. K. (2009). Use of patient-reported outcomes in clinical practice. *Lancet*, *374*(9687), 369–370.
- OECD. (2019). Measuring what matters the Patient Reported Indicator Surveys.
- ePROVIDE. (2021). Databases on Clinical Outcome Assessments (COAs). Retrieved 05/07/2021, from <https://eprovide.mapi-trust.org/>
- Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., Filiberti, A., Flechtner, H., Fleishman, S. B., de Haes, J. C., & Kaasa, S. (1993). 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*, *85*(5), 365–376.
- Cella, D., Choi, S. W., Condon, D. M., Schalet, B., Hays, R. D., Rothrock, N. E., Yount, S., Cook, K. F., Gershon, R. C., Amtmann, D., DeWalt, D. A., Pilkonis, P. A., Stone, A. A., Weinfurt, K., & Reeve, B. B. (2019). PROMIS® adult health profiles: efficient short-form measures of seven health domains. *Value Health*, *22*(5), 537–544.
- Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., Amtmann, D., Bode, R., Buysse, D., Choi, S., Cook, K., Devellis, R., DeWalt, D., Fries, J. F., Gershon, R., Hahn, E. A., Lai, J.-S., Pilkonis, P., Revicki, D., ... Group P. C. (2010). The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of clinical epidemiology*, *63*(11), 1179–1194.
- Jr Ware, J. E. (1984). Methodology in behavioral and psychosocial cancer research. Conceptualizing disease impact and treatment outcomes. *Cancer*, *53*(10 Suppl), 2316–2326.
- Ware, J. E., Brook, R. H., Davies, A. R., Williams, K. N., Stewart, A., Rogers, W. H., Donald, C. A., & Johnston, S. A. (1980). Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. I, Model of Health and Methodology: RAND Corporation.
- Pearman, T. P., Beaumont, J. L., Cella, D., Neary, M. P., & Yao, J. (2016). Health-related quality of life in patients with neuroendocrine tumors: An investigation of treatment type, disease status, and symptom burden. *Supportive Care in Cancer*, *24*(9), 3695–3703.
- Henry, N. L., Kidwell, K. M., Alsamraie, C., Bridges, C. M., Kwiatkowski, C., Clauw, D. J., Smith, E. M. L., & Williams, D. A. (2018). Pilot study of an internet-based self-management program for symptom control in patients with early-stage breast cancer. *JCO Clinical Cancer Informatics*, *2*, 1–12.
- Halperin, D. M., Huynh, L., Beaumont, J. L., Cai, B., Bhak, R. H., Narkhede, S., Totev, T., Duh, M. S., Neary, M. P., & Cella, D. (2019). Assessment of change in quality of life, carcinoid syndrome symptoms and healthcare resource utilization in patients with carcinoid syndrome. *BMC Cancer*, *19*(1), 274.
- Snyder, C. F., Herman, J. M., White, S. M., Luber, B. S., Blackford, A. L., Carducci, M. A., & Wu, A. W. (2014). When using patient-reported outcomes in clinical practice, the measure matters: A randomized controlled trial. *Journal of Oncology Practice*, *10*(5), e299–e306.
- Cella, D., Gershon, R., Lai, J. S., & Choi, S. (2007). The future of outcomes measurement: Item banking, tailored short-forms, and computerized adaptive assessment. *Quality of Life Research*, *16*(Suppl 1), 133–141.
- Liegl, G., Petersen, M. A., Groenvold, M., Aaronson, N. K., Costantini, A., Fayers, P. M., Holzner, B., Johnson, C. D., Kemmler, G., Tomaszewski, K. A., Waldmann, A., Young, T. E., Rose, M., & Nolte, S. (2019). Establishing the european norm for the health-related quality of life domains of the computer-adaptive test EORTC CAT Core. *European Journal of Cancer*, *107*, 133–141.
- Fischer, F., Gibbons, C., Coste, J., Valderas, J. M., Rose, M., & Leplège, A. (2018). Measurement invariance and general population reference values of the PROMIS Profile 29 in the UK, France, and Germany. *Quality of Life Research*, *27*(4), 999–1014.

19. Cella, D., Choi, S. W., Condon, D. M., Schalet, B., Hays, R. D., Rothrock, N. E., Yount, S., Cook, K. F., Gershon, R. C., Amtmann, D., DeWalt, D. A., Pilkonis, P. A., Stone, A. A., Weinfurt, K., & Reeve, B. B. (2019). PROMIS® adult health profiles: Efficient short-form measures of seven health domains. *Value Health*, 22(5), 537–544.
20. Fried, E. I. E. d. B. (2021). From mandating common measures to mandating common metrics: A plea to harmonize measurement results. PsyArXiv.
21. U.S. Food and drug administration. (2021). Core patient-reported outcomes in cancer clinical trials.
22. EORTC. QUESTIONNAIRES : CORE. Retrieved 05/07/2021, from <https://qol.eortc.org/core/>
23. Wechsler, S. (1997). *Statistics at Square One*. Ninth Edition, revised by M. J. Campbell, T. D. V. Swinscow, BMJ Publ. Group, London, 1996. No. of pages: 140. Price: £11. ISBN 0-7279-0916-9. *Statistics in Medicine*. 16(22): 2629-2630
24. Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
25. Hays, R. D., Spritzer, K. L., Schalet, B. D., & Cella, D. (2018). PROMIS®-29 v2.0 profile physical and mental health summary scores. *Quality of Life Research*, 27(7), 1885–1891.
26. Costa, D. S., Aaronson, N. K., Fayers, P. M., Grimison, P. S., Janda, M., Pallant, J. F., Rowen, D., Velikova, G., Viney, R., Young, T. A., & King, M. T. (2014). Deriving a preference-based utility measure for cancer patients from the european organisation for the research and treatment of cancer's quality of life questionnaire C30: A confirmatory versus exploratory approach. *Patient Related Outcome Measures*, 5, 119–129.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Komplette Publikationsliste

Peer-reviewed Journals

- (1) **Hartmann C.**, Fischer F., Klapproth C.P., Röhle R., Rose M. & Karsten M.M. PROMIS-29 and EORTC QLQ-C30: an empirical investigation towards a common conception of health. *Qual Life Res.* 2023;32(3):749-758.
- (2) Plessen, C. Y., Liegl, G., **Hartmann, C.**, Heng, M., Joeris, A., Kaat, A. J., Schalet, B. D., Fischer, F. & Rose, M. (2023). How Are Age, Gender, and Country Differences Associated With PROMIS Physical Function, Upper Extremity, and Pain Interference Scores? *Clinical Orthopaedics And Related Research.* <https://doi.org/10.1097/corr.0000000000002798>
- (3) **Hartmann C.**, Rose M., Weichert A. & Weißhaupt K. The ICHOM Standard Set for Pregnancy and Childbirth - Translation and Linguistic Adaptation for Germany. *Geburtsh Frauenheilk.* 2022;82(7):747-754.
- (4) Klapproth, C. P., **Hartmann, C.**, Fischer, F. & Rose, M. (2022). POSC341 Comparison of PROPr and EQ-5D-3L in Lower Back Pain Patients. *Value in Health*, 25(1), S238. <https://doi.org/10.1016/j.jval.2021.11.1166>
- (5) Maechler F., Gertler M., Hermes J., van Loon W., Schwab F., Piening B., Rojansky S., Hommes F., Kausch F., Lindner A.K., Burock S., Rössig H., **Hartmann C.**, Kirchberger V., Thieme A., Behnke M., Gastmeier P., Mockenhaupt F.P. & Seybold J. Epidemiological and clinical characteristics of SARS-CoV-2 infections at a testing site in Berlin, Germany, March and April 2020—a cross-sectional study. *Clin Microbiol Infect.* 2020;26(12):1685.e7-1685.e12.
- (6) Karsten M.M., Speiser D., **Hartmann C.**, Zeuschner N., Lippold K., Kiver V., Gocke P., Kirchberger V. & Blohmer J.U. Web-Based Patient-Reported Outcomes Using the International Consortium for Health Outcome Measurement Dataset in a Major German University Hospital: Observational Study. *JMIR Cancer.* 2018;4(2):e11373.

Andere Puplikationen:

- (1) **Hartmann C.**, Jaha H., Hofmann L., Schappert A., Rose M. Messung von Lebensqualität in der Nephrologie. *Nephrologie Aktuell*, 28(02), 79–84.
- (2) **Hartmann C.** & Fischer F. Die Erhebung von Patient-Reported Outcomes In: *Das digitale Krankenhaus*. 1.Aufl. Medizinisch Wissenschaftliche Verlagsgesellschaft (MWV). 2023:359-364.
- (3) **Hartmann C.**, Obbarius A., Haneke H., Paul Klapproth C. & Rose M. Value-based Healthcare - Gesundheitsversorgung neu denken *OP-Management up2date*. 2021;1(1):87-105.
- (4) **Hartmann C.**, Fischer K., Haneke H. & Kirchberger V. Patient-reported outcomes in spinal surgery—how can we keep getting better? *J Spine Surg.* 2020;6(4):820-824.

Danksagung

Zunächst möchte ich mich bei Prof. Matthias Rose für die Möglichkeit bedanken, am Center for Patient-Centered Outcome Research (CPCOR) zu promovieren und mir neben der praktischen Arbeit zu Patient-Reported Outcomes auch relevantes Fachwissen anzueignen. Vor allem aber danke ich Ihm dafür, meinen Forschergeist geweckt zu haben.

Ein besonderer Dank geht an Dr. med. Valerie Kirchberger, ohne die sich die Chance einer Promotion nie für mich eröffnet hätte. Danke, dass du an mich geglaubt hast und dein stets offenes Ohr.

Mein Dank gilt außerdem einer Reihe von Menschen, die wesentlich zum Gelingen der Arbeit beigetragen haben. Dem gesamten Team des CPCOR für die stets offene und unterstützende Atmosphäre. Ich schätze den intensiven fachlichen aber auch persönlichen Austausch mit Euch sehr. Meiner Freundin und Kollegin Norma Bethke für den Zuspruch, die Unterstützung und den engen Austausch in den Wochen vor der finalen Einreichung. Ein großer Dank gilt auch Frederike Peters für ihre wertvollen Anregungen, ihre Geduld und ihr offenes Ohr in dieser intensiven Zeit.

Meiner Familie und meinem Partner danke ich von Herzen für die mentale Unterstützung, die große Wertschätzung, die ihr mir und meiner Arbeit entgegenbringt, sowie für das Vertrauen in mich und meine Fähigkeiten.